TUOMAS KILPELÄINEN

Population-Based Screening for Prostate Cancer

Weighing up the beneficial and adverse effects of screening

■

ACADEMIC DISSERTATION
To be presented, with the permission of
the board of the School of Medicine of the University of Tampere,
for public discussion in the Small Auditorium of Building M,
Pirkanmaa Hospital District, Teiskontie 35,
Tampere, on November 30th, 2012, at 12 o'clock.

UNIVERSITY OF TAMPERE

UNIVERSITY
OF TAMPERE

ACADEMIC DISSERTATION
University of Tampere, School of Medicine
University of Tampere, School of Health Sciences
Finland

Supervised by
Professor Teuvo Tammela
University of Tampere
Finland
Professor Anssi Auvinen
University of Tampere
Finland

Reviewed by
Docent Antti Rannikko
University of Helsinki
Finland
Professor Gunnar Steineck
University of Gothenburg
Sweden

Distribution
Bookshop TAJU
P.O. Box 617
33014 University of Tampere
Finland

Tel. +358 40 190 9800
Fax +358 3 3551 7685
taju@uta.fi
www.uta.fi/taju
http://granum.uta.fi

Cover design by
Mikko Reinikka

Acta Universitatis Tamperensis 1782
ISBN978-951-44-8973-0 (print)
ISSN-L 1455-1616
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 1257
ISBN 978-951-44-8974-7 (pdf)
ISSN 1456-954X
http://acta.uta.fi

Tampereen Yliopistopaino Oy – Juvenes Print
Tampere 2012

Contents

List of original contributions

I Kilpeläinen TP, Tammela TL, Määttänen L, Kujala P, Stenman UH, Ala-Opas M, Murtola TJ, Auvinen A. False-positive screening results in the Finnish prostate cancer screening trial. Br J Cancer 2010;102:469-74.

II Kilpeläinen TP, Tammela TL, Roobol M, Hugosson J, Ciatto S, Nelen V, Moss S, Määttänen L, Auvinen A. False-positive screening results in the European randomized study of screening for prostate cancer. Eur J Cancer 2011;47:2698-705.

III Kilpeläinen TP, Auvinen A, Määttänen L, Kujala P, Ruutu M, Stenman UH, Tammela TL. Results of the three rounds of the Finnish Prostate Cancer Screening Trial – the incidence of advanced cancer is decreased by screening. Int J Cancer 2010;127:1699-705.

IV Kilpeläinen TP, Tammela TL, Malila N, Hakama M, Santti H, Määttänen L, Stenman UH, Kujala P, Auvinen A. Prostate cancer mortality in the Finnish screening trial. [Submitted.]

Abbreviations

ASAP	Atypical small acinar proliferation
BPH	Benign prostatic hyperplasia
CI	Confidence interval
DRE	Digital rectal examination
ERSPC	European Randomized Study of Screening for Prostate Cancer
F/T PSA	Free/total PSA ratio
FP	False-positive (screening result)
HG-PIN	High-grade prostatic intraepithelial neoplasia
HR	Hazard ratio
IRR	Incidence rate ratio
NNS	Number needed to screen
OR	Odds ratio
PC	Prostate cancer
PCPT	Prostate Cancer Prevention Trial
PIN	Prostatic intraepithelial neoplasia
PLCO	Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial
PPV	Positive predictive value
PSA	Prostate-specific antigen
REDUCE	Reduction by Dutasteride of Prostate Cancer Events
RR	Risk ratio (or rate ratio)
TRUS	Transrectal ultrasound

Abstract

Prostate cancer (PC) has become a major challenge for public health since a steep rise in incidence was observed in the 1990s. Screening for PC with prostate-specific antigen (PSA) has been a controversial issue for two decades, and reliable evidence for or against screening can be obtained only from large randomized trials. The European Randomized Study of Screening for Prostate Cancer (ERSPC) was commenced in the early 1990s as an international multicenter trial to assess the benefits and harms of screening for PC with PSA. The Finnish prostate cancer screening trial is the largest component of the ERSPC trial. The aim of this dissertation was to assess the prevalence of false-positive (FP) screening results both in the Finnish trial and in four other ERSPC centers, and to determine how screening affected the incidence of PC and mortality in the Finnish trial.

The Finnish population-based trial included 80,144 men. The men were randomized either to the screening arm (SA) or the control arm (CA). The men in the CA received the usual care and were not contacted. The men in the SA were tested with PSA at four-year intervals up to three times. Men with PSA ≥4.0 ng/ml were referred to diagnostic work-up (prostate biopsies), as were those with PSA 3.0-3.99 ng/ml and positive ancillary test (digital rectal examination in 1996-1998; free/total PSA ratio <16% from 1999 onwards).

In Finland, 75% of the men in the SA participated at least once in screening, and the participation proportion was roughly 70% per round. In multi-center analyses, on average every tenth man attending screening was subject to an FP screening result. A quarter of men with an FP result chose not to participate in the next screening round, and of the men who did participate, approximately 50% had another FP result. The risk for a next-round PC was roughly four-fold compared to men with a previous negative screen.

In the Finnish trial, the overall incidence of PC was 8.5/1,000 person-years in the SA and 6.3/1,000 person-years in the CA (incidence rate ratio (IRR) = 1.30; 95% confidence intervals 1.24-1.36). The IRR for localized PC was 1.48 (1.40-1.56) and for advanced PC 0.74 (0.66-0.83). The cumulative mortality was 59/10,000 men in the SA and 65/10,000 men in the CA [hazard ratio (HR) 0.89 (0.75-1.07)] during 12 years of follow-up. After adjusting for bias caused by non-participation, the ratio was only slightly improved. According to a secondary analysis, a lower PSA threshold of 3.0 ng/ml would not have improved the HR substantially, whereas excluding all interval cancers would have yielded a more prominent decrease in mortality in the SA.

Screening for PC with PSA resulted in a substantial (30%) decrease in the incidence of advanced PC, but this did not transform into a statistically significant decrease in PC mortality during 12 years of follow-up. The men screened were often subject to an FP screening result and overdiagnosed cancers. The results of this dissertation do not support launching generalized population-based PC screening. It is possible that with longer follow-up the mortality reduction becomes larger, but even then cost-effectiveness and quality of life factors need to be weighed against the observed benefits.

Tiivistelmä

Eturauhassyöpä on teollistuneissa maissa miesten yleisin syöpä. Prostata-spesifistä antigeeniä (PSA) on käytetty eturauhassyövän diagnostiikassa jo vuosikymmeniä, mutta sen käytöstä eturauhassyövän seulonnassa ei kuitenkaan ole tieteellistä näyttöä. European Randomized Study of Screening for Prostate Cancer (ERSPC) on kansainvälinen, 90-luvun alussa käynnistetty prospektiivinen randomisoitu monikeskustutkimus, jonka yhtenä päätarkoituksena on selvittää, voidaanko PSA-seulonnalla vähentää eturauhassyöpäkuolleisuutta. Suomen väestöpohjainen eturauhassyöpä-seulontatutkimus on osa ERSPC-tutkimusta. Tämän väitöskirjan tavoitteena oli selvittää, kuinka yleisiä väärät positiiviset seulontatulokset ovat paitsi Suomen seulontatutkimuksessa, myös neljässä muussa ERSPC-keskuksen aineistossa. Lisäksi tavoitteena oli arvioida, miten PSA-seulonta vaikuttaa eturauhassyövän ilmaantuvuuteen ja kuolleisuuteen Suomessa.

Suomen seulontatutkimuksessa oli yli 80 000 miestä, joista noin 32 000 satunnaistettiin seulontaryhmään ja loput muodostivat verrokkiryhmän. Seulontaryhmän miehet kutsuttiin neljän vuoden välein PSA-testiin, ja seulontakierroksia oli yhteensä kolme. Seulottu mies lähetettiin jatkotutkimuksiin (eturauhaskoepalat), mikäli PSA-taso oli ≥4,0 ng/ml, tai mikäli PSA-taso oli 3,0 – 3,99 ng/ml jos lisätesti oli poikkeava (epäilyttävä tuseerauslöydös vuosina 1996 – 1998; vapaan PSA:n suhde kokonais-PSA:iin <16 % vuodesta 1999 lähtien).

Suomessa seulontaryhmän miehistä 75 % osallistui ainakin kerran seulontaan, ja osallistumisprosentti yhtä kierrosta kohden oli keskimäärin 70 %. ERSPC-keskusten yhteistuloksissa väärän positiivisen seulontatuloksen sai noin joka kymmenes seulontaan osallistuneista. Väärän positiivisen tuloksen saaneista noin neljännes ei osallistunut seuraavalle kierrokselle, ja yli puolet

niistä, jotka osallistuivat, saivat uudelleen väärän positiivisen seulontatuloksen. Väärän positiivisen tuloksen saaneilla oli noin nelinkertainen riski saada seuraavalla osallistumiskerralla eturauhassyöpädiagnoosi seulontanegatiivisiin verrattuna.

Suomen aineistossa eturauhassyövän ilmaantuvuus oli seulontaryhmässä 8,5 (1 000 henkilövuotta kohden) ja verrokkiryhmässä 6,3 (ilmaantuvuussuhde (IRR) 1,30; 95 % luottamusvälit 1,24–1,36). Paikallisen syövän IRR oli 1,48 (1,40–1,56) ja levinneen syövän 0,74 (0,66–0,83). Eturauhassyövän kertymäkuolleisuus oli seulontaryhmässä 59 (10 000 miestä kohden) ja verrokkiryhmässä 65 (vaarasuhde 0,89 (0,75–1,07). Kun osallistumattomien miesten aiheuttama harha korjattiin lisäanalyyseissä, ei merkittävää parannusta tapahtunut. Myöskään matalamman PSA-rajan (3,0 ng/ml) käyttäminen ei olisi lisäanalyysien mukaan vaikuttanut kuolleisuuteen, mutta ns. välisyöpien (seulontakierrosten välissä ilmaantuneiden syöpien) välttäminen olisi parantanut kuolleisuustulosta merkitsevästi.

Eturauhassyövän seulonta vähensi levinneen syövän ilmaantuvuutta noin kolmanneksella. Tämä ei kuitenkaan johtanut tilastollisesti merkitsevään kuolleisuusalenemaan noin 12 vuoden seuranta-ajalla. Seulonnasta aiheutui tutkituille vääriä positiivisia tuloksia ja ylidiagnosoituja eturauhassyöpiä. On mahdollista, että pidemmän seuranta-ajan kuluessa kuolleisuusero kasvaa, mutta tässäkin tapauksessa seulonnan haitat ylidiagnostiikan, elämänlaatukysymysten ja kustannusvaikuttavuuden suhteen on selvitettävä ennen kuin väestöpohjaista seulontaa voidaan suositella.

1. INTRODUCTION

After decades of research and debate, prostate cancer (PC) screening is still a controversial issue. Since prostate-specific antigen (PSA) was identified in the 1980s as a risk marker for PC, great interest has been focused on using PSA in mass screening for PC. PC is a major public health issue, as in Finland, for example, the cumulative risk for being diagnosed with it is 11% before the age of 75, and 800 men die of PC annually [Engholm et al. 2011].

In 2009 the European Randomized Study of Screening for Prostate Cancer (ERSPC) provided evidence for the first time that PC mortality can be decreased with screening [Schröder et al. 2009]. This relative reduction could not be seen in another large prospective randomized trial, the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) in the USA [Andriole et al. 2009]. The contradictory results from these trials have added fuel to the discussion regarding PC screening. Even if PC screening reduced mortality, several questions remain unanswered: Does screening lead to improved quality of life? Do the benefits of screening outweigh the harms caused by screening? Is screening cost-effective? Before these questions are answered, no decision to commence nationwide screening can be taken.

This dissertation was based on the Finnish component of the ERSPC trial. It set out to investigate how screening affects PC incidence and mortality in Finland, and to study a common adverse effect, false-positive (FP) screening results both in the Finnish trial and in four other ERSPC centers.

2. REVIEW OF THE LITERATURE

2.1 Anatomy and function of the prostate

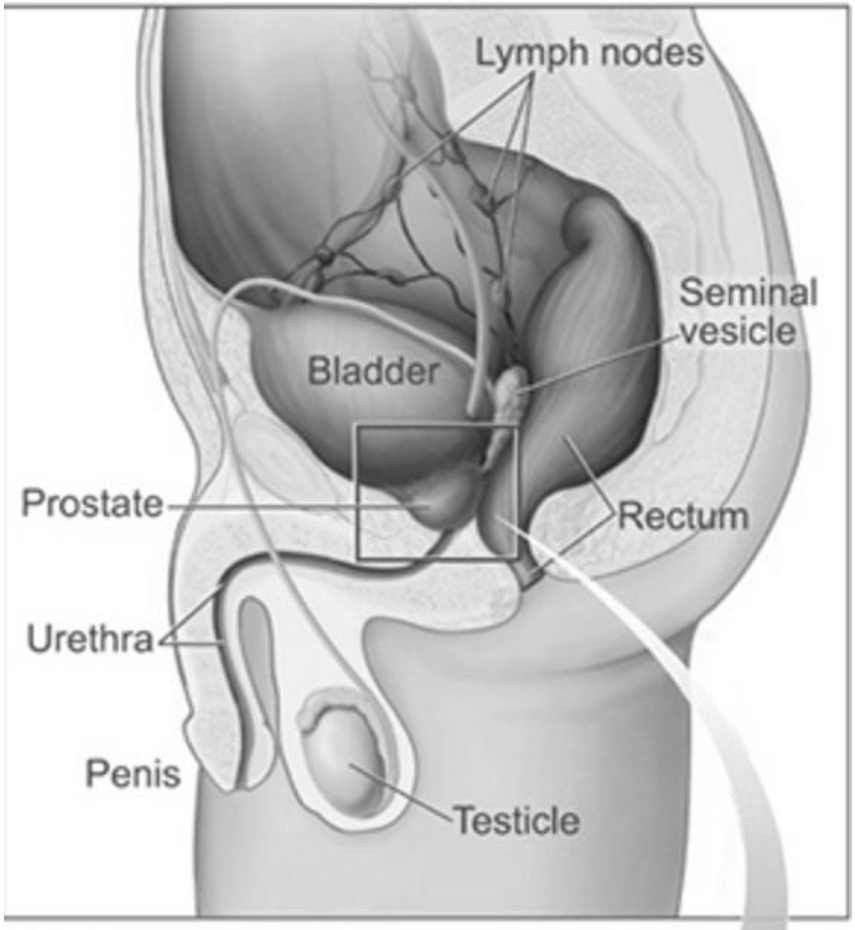
The prostate (Greek: *prostates*; "the one who stands before and protects") is the largest accessory gland of the male reproductive system. It is located in the pelvis, caudad of the urinary bladder, surrounding the proximal urethra (Figure 2.1). Approximately two thirds of the prostate is glandular tissue, one third consists of fibromuscular tissue. The prostate can be divided into peripheral and central zones; the first is a common area for carcinomas and the latter is usually involved in benign prostatic hyperplasia (BPH). In an adult man, the prostate measures approximately 3 cm and weighs ca. 11 g [Leissner and Tisell 1979].

The function of the prostate is to secrete prostatic fluid that constitutes one fifth of the volume of the semen. This milky, thin fluid is slightly alkaline, and contains mainly acid phosphatase, fibrinolysin, PSA, zinc, and citric acid. The fluid is stored in the prostate and ejected to the urethra at the time of ejaculation. The function and growth of the prostatic cells are driven by androgens, namely dihydrotestosterone, which is converted by an enzyme 5α-reductase from testosterone (which in turn is produced by the testes).

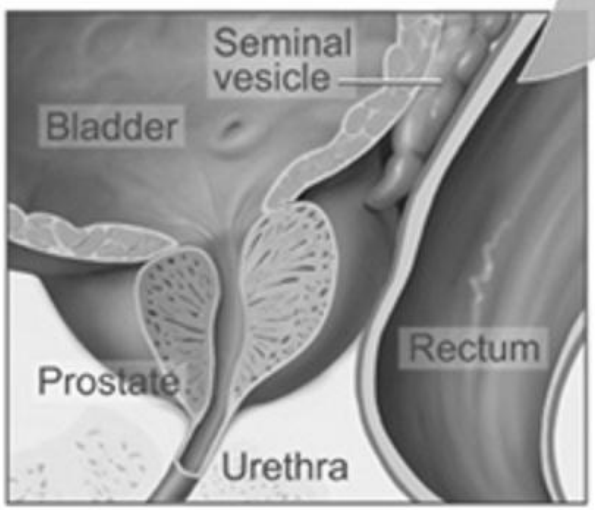
2.2 Benign prostatic disorders

Prostate cancer is a malignant prostatic disorder, but several benign conditions may afflict the prostate. The prostate may become enlarged due to benign prostatic hyperplasia (BPH) – a condition which becomes more common with age [Koskimäki et al. 1998; Rhodes et al. 1999]. An enlarged prostate may compress the proximal urethra and impede urinary flow, causing lower urinary

Figure 2.1. The prostate is situated caudad of the urinary bladder, surrounding the proximal urethra, anterior to the rectum. [Image downloaded from Wikimedia Commons, commons.wikimedia.org.]



This shows the prostate and nearby organs.



This shows the inside of the prostate, urethra, rectum, and bladder.

tract symptoms such as nocturia, difficulty to void, frequent need to urinate, incontinence and sometimes urinary retention.

BPH may be treated conservatively with medications such as alpha blockers (which relax the smooth muscle in the prostate and bladder neck) or 5α-reductase inhibitors (which inhibit the proliferative effect of dihydrotestosterone on the prostatic cells). In some cases surgical treatment for BPH is indicated: transurethral incision or resection of the prostate may be chosen, likewise thermotherapy, needle ablation or newer laser techniques.

So far, there is no conclusive evidence that BPH constitutes a risk factor or is a prerequisite for PC, although BPH is often present in prostates with carcinoma, both conditions are fueled by androgens and both conditions become more common with age [Alcaraz et al. 2009; Bushman 2009].

Prostatitis, an inflammation of the prostate may be acute or chronic, bacterial or nonbacterial. Prostatitis causes similar lower urinary tract symptoms as BPH, but in addition often causes pain. The onset of symptoms may be sudden (in acute bacterial prostatitis) or intermittent (in nonbacterial chronic prostatitis, which is also known as chronic pelvic pain syndrome).

Chronic inflammation has been postulated to increase the risk for PC, as inflammation causes recurring cell damage through oxidative stress and promotes cell proliferation. A five-year follow-up study found an increased risk for carcinoma at rebiopsy in men who had inflammatory changes in their prostate biopsy samples five years earlier (20% (N=29/144; 95% CI 14-28%) vs. 6% (N=2/33; 95% CI 0.7-20%)) [MacLennan et al. 2006]. Due to the relatively low number of observations in this study the confidence intervals (not presented in the original study) for aforementioned proportions overlapped and thus remained inconclusive. In the screening setting, prostate biopsies with inflammation did not predict diagnosis of PC four years later [Wolters et al. 2008]. A meta-analysis by Dennis et al., on the other hand, did report a borderline significantly increased risk for PC in men with history of prostatitis (OR = 1.6, 95% CI 1.0-2.4), particularly in population-based case-control studies (OR = 1.8, 95% CI 1.1-3.0) [Dennis et al. 2002]. The role of inflammation in the development of BPH and PC needs further elucidation [Alcaraz et al. 2009].

Several pathological lesions can be found in the prostate. Atypical adenomatous hyperplasia is a benign proliferative acinar lesion mimicking the appearance of a low-grade adenocarcinoma. Its role in prostatic tumorigenesis has been unclear, but accumulating evidence suggests that atypical adenomatous hyperplasia is not a precursor to prostatic carcinoma [Meyer et al. 2006; Epstein 2009]. An atypical small acinar proliferation (ASAP; also known as 'atypical lesion suspicious for PC') is also an acinar proliferation, but differs from the benign atypical adenomatous hyperplasia in that ASAP has some features suspicious of carcinoma, but not sufficient for the diagnosis of carcinoma. The point prevalence of ASAP in asymptomatic men in a multicenter screening trial has been estimated at 2.2-2.7%; furthermore, at rebiopsy 2-4 years later 34% of these men were diagnosed with prostatic carcinoma [Laurila et al. 2010].

Prostatic intraepithelial neoplasia (PIN) consists of architecturally benign acini but cytologically atypical cells lining the acini, and is often found in the same peripheral zone from which prostate carcinomas commonly originate. Two grades of PIN have been identified: low-grade and high-grade (HG-PIN), but interobserver variability is high in low-grade PINs, which limits its clinical use [Bostwick et al. 2004]. Low-grade PIN does not appear to increase the risk for subsequent PC [Vis and van der Kwast 2001]. HG-PIN however, is more consistently acknowledged among pathologists [Epstein 2009]. The risk for PC after HG-PIN in biopsy has been assessed in several studies, with considerable range: 2.3% to 100%; median 24% [Epstein and Herawi 2006]. Notably the higher proportions were recorded in the era preceding the widespread use of PSA in asymptomatic men. More recent studies have demonstrated that the positive predictive value (PPV) of HG-PIN has decreased over the years resembling PPV after a benign biopsy, and that HG-PIN is not necessarily an indication for early rebiopsy [Vis and van der Kwast 2001; Postma et al. 2004; Schlesinger et al. 2005; Laurila et al. 2010].

Finally, prostatic atrophy (especially proliferative inflammatory atrophy) has been proposed as a precursor to both HG-PIN and carcinoma [De Marzo et al. 2004]. This notion is supported by the fact that in radical prostatectomy samples there is a spatial relationship with proliferative inflammatory atrophy

and carcinoma foci [Putzi and De Marzo, 2000; Wang et al. 2009]. Furthermore, there are similar molecular changes in atrophy, HG-PIN and carcinoma [Palapattu et al. 2005]. However, not all HG-PIN or carcinoma lesions are associated with atrophy [Putzi and De Marzo, 2000], and atrophy is moreover a very common finding in the prostates of aging men: in a screening trial, atrophy was found in 94% of prostate biopsies [Postma et al. 2005]. In the same study, the incidence of PC was not increased after atrophy was found in biopsy (follow-up of eight years). Another type of atrophy, sclerotic atrophy, is quite rare (9% of the biopsies in the study by Postma et al.) and has not been proposed as a risk factor for PC.

2.3 Prostate cancer

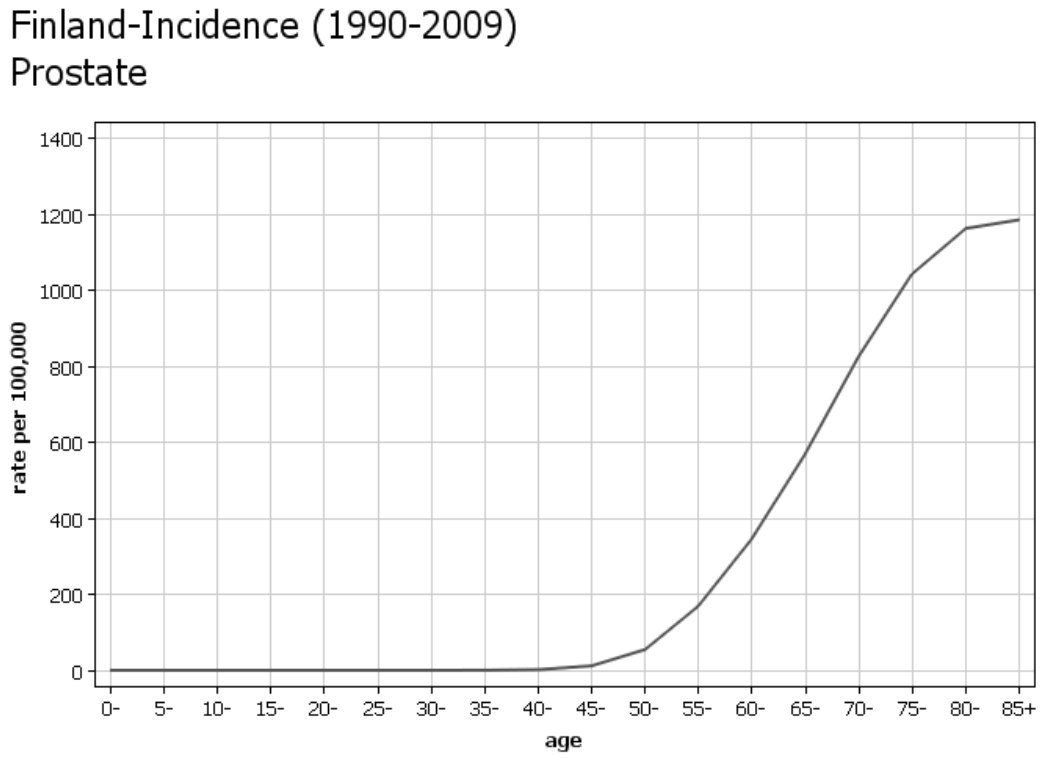
2.3.1 Etiology

2.3.1.1 Age and ethnicity

Prostate cancer risk increases markedly with age, and the disease seldom affects men under 50 years of age (Figure 2.2). For example, in Finland in 1990-2009, the incidence rate for men in the age group 50-54 years was 55/100,000 person-years, but the incidence rate increased rapidly and was 1042/100,000 person-years in the age group 75-79 years [Engholm et al. 2011].

In addition to age, another well-established risk factor for PC is ethnicity. According to the data from the Surveillance Epidemiology and End Results in the USA, the PC incidence in 2002-2006 was 153/100,000 person-years in white males and 240/100,000 person-years in black males [Horner et al. 2009]. In a study reviewing over 12,000 men with PC, African American men were also more likely to be diagnosed with PC before the age of 50 than were Caucasian men (8.3% vs. 3.3%, $p < 0.0001$) [Parker et al. 2011]. A recent review article by Mordukhovich et al. pooled 37 articles focusing on the elevated PC risk in African American males compared to Caucasian males in the USA, but found no

Figure 2.2. Incidence of prostate cancer in Finland in 1990-2009 increases rapidly after the age of 50 years. [Engholm et al. 2011]



conclusive difference in environmental risk factors nor in family history of PC that could explain the marked difference between the two ethnic groups [Mordukhovich et al. 2011].

Interestingly, the risk for PC in Japanese men increases after immigration to the USA, from 8/100,000 to 32/100,000 [Shimizu et al. 1991]. The incidence still remains lower than in men of Japanese ethnicity born in the USA – namely 30/100,000 vs. 42/100,000 [Cook et al. 1999].

There is much variation in the incidence of PC worldwide (discussed in detail in Chapter 2.4). However, there is evidence that the frequency of microscopic foci of prostate carcinoma does not vary as much: in all seven areas studied – Hong Kong, Singapore, Sweden, the former Federal Republic of Germany, Jamaica, Uganda and Israel – the frequency of these small latent carcinomas was approximately 12% [Breslow et al. 1977]. In other autopsy studies, these microscopic foci of carcinoma were found in as many as 15-29% of men aged

30-40 years, and in 50-64% of men aged 60-70 years [Sakr et al. 1994; Soos et al. 2005].

In light of the aforementioned studies, it is evident that small, latent foci of prostate carcinoma are common even in relatively young men, and these foci are found as often in countries with high or low incidence of clinical PC. The progression from relatively common latent PC to clinically significant disease remains a mystery and is likely to involve interplay between genetic and environmental issues.

2.3.1.2 Hereditary factors

Besides age and ethnicity, the third well-established risk factor for PC is family history. In the 1950s, men with a first-degree relative affected by PC were found to be at increased risk for PC themselves [Morganti et al. 1956-1957]. Since then, several case-control and population-based cohort or registry studies have confirmed the finding. Three comprehensive meta-analyses published in 2003 reviewed the earlier publications [Bruner et al. 2003; Johns and Houlston 2003; Zeegers et al. 2003]. Bruner et al. analyzed 24 studies and found that in men with an affected father the risk ratio (RR) was 2.1 (95% CI 1.8-2.5), whereas in men with an affected brother the RR was somewhat higher, 2.9 (95% CI 2.2-3.7). Johns and Houlston reviewed 13 case-control and cohort studies and resulted in RR of 2.5 (95% CI 2.2-2.8) for men with a first-degree relative diagnosed with PC; this risk was even higher for men with two affected relatives: RR 3.5 (95% CI 2.6-4.8). Zeegers et al. analyzed 33 studies and reported a RR of 2.5 (95% CI 2.2-2.9). In all three meta-analyses the RR was higher for brothers of cases compared to the sons of cases. This may reflect the fact that brothers are more likely to share similar environmental and behavioristic factors in the course of their lives than with their fathers [Bruner et al. 2003].

Hereditary factors appear to be significant, especially in men diagnosed with PC at an early age (<50 years) [Lesko et al. 1996; Matikainen et al. 2001; Johns and Houlston 2003; Zeegers et al. 2003]. The increased risk remains even after controlling for environmental factors (smoking, alcohol consumption, body

mass index, physical activity, education, history of sexually transmitted diseases, diet, and hormone levels) [Kalish et al. 2000]. Men with an affected first-degree relative are also at increased risk for dying of PC (risk ratio (RR) = 1.6; 95% CI 1.3-2.0), particularly if they had two or more affected first-degree relatives (RR 3.2; 95% CI 1.5-6.7) [Rodríguez et al. 1997]. The increased risk for familial PC appears to be similar in Caucasians, African Americans and Asian Americans [Hayes et al. 1995; Whittemore et al. 1995].

It has previously been estimated that 5-10% of all PCs have a strong genetic component [Carter et al. 1992; Bratt et al. 1999; Cerhan et al. 1999]. However, an analysis utilizing the Swedish Cancer Registry went through 34 cancer sites in a population <72 years of age and discovered that PC showed the highest familial tendency, 20% [Hemminki et al. 2008]. Another study analyzing Nordic twins estimated that 16-45% of PCs showed genetic susceptibility (compared to 12-30% in breast cancer and 8-27% in colorectal cancer) [Baker et al. 2005].

Furthermore, as many as 30-40% of early onset PCs (diagnosed before the age 55 years) may be attributable to hereditary factors [Carter et al. 1993; Bratt et al. 1999]. Hereditary PC typically has 6-7 years earlier onset of disease, but there is no conclusive evidence that these cancers are otherwise clinically different from sporadic cancers [Norrish et al. 1999; Valeri et al. 2000; Bratt et al. 2002].

The level of circulating androgens (namely testosterone and/or its active metabolite dihydrotestosterone) has also been suggested to affect the risk for PC, since these fuel both normal prostatic cells and prostatic cancer cells. According to two meta-analyses, endogenous hormone levels are not associated with PC [Eaton et al. 1999; Roddam et al. 2008].

So far no major predisposing gene has been found in PC [Colloca and Venturino 2011]. Several susceptibility loci – found in every human chromosome – have been suggested to be associated with PC [Christensen et al. 2010; Ioannidis et al. 2010], but finding conclusive evidence has proven challenging [Ostrander and Stanford 2000]. Firstly, PC is a very common disease, and thus there may be also sporadic cases also in those non-afflicted members of the families with genetic components predisposing to PC. Secondly, as previously mentioned, the onset age of hereditary PC is relatively close to

that of sporadic cases compared to e.g. breast and colorectal cancer (for which the onset of hereditary cancers may be advanced by 20 years), which makes identifying hereditary PCs from sporadic ones difficult. Thirdly, since there is no single major predisposing gene (as for example, in breast cancer BRCA1 and BRCA2), but instead several polymorphisms that in different populations moderately increase the PC risk [Christensen et al. 2010], it is likely that the overall risk for PC is driven by a complex multitude of genetic factors. Novel genome-wide association studies have identified several susceptibility loci, and probably will find more in the future. However, it has been estimated that the susceptibility loci that have not yet been found are likely to be of even less importance in prostatic tumorigenesis than those already identified [Ioannidis et al. 2010]. In addition, the interplay with environmental factors renders finding associations even more complex.

In general, studying hereditary PC is challenging, as family history is not only inherited genes, but also inherited behavior and similar environmental factors. The men who have affected men in their family may be more likely to seek urological advice even if they do not have any urological symptoms, which – as PC is so common – could result in overdiagnosis (detection bias) and subsequently, lead time bias (discussed in Chapter 2.5).

2.3.1.3 Environmental factors

Given that PC incidence varies so much between geographical areas (see Section 2.4), hypotheses regarding environmental risk factors have been extensively pursued. Of the traditional risk factors for cancer, alcohol is generally not regarded as a risk factor for PC. A large European cohort study found no statistically significant association between alcohol and PC [Rohrmann et al. 2008]. Also, a meta-analysis in 2000 concluded that although there are some studies showing an association, these are likely to be biased towards finding a positive association – thus there is no conclusive evidence associating alcohol and increased risk for PC (RR 1.05; 95% CI 0.98-1.11) [Dennis 2000]. For another well-known carcinogen, smoking, the evidence is also inconclusive. Smoking does not seem to increase the risk for developing PC, as evinced by two

comprehensive review articles [Hickey et al. 2001; Zu et al. 2009]. Smokers do, on the other hand, appear to have more advanced disease at diagnosis, poorer prognosis and a greater risk for fatal PC [Zu et al. 2009]. Estimating the effect of smoking, however, is difficult since smoking is often associated with lower socio-economic status and co-morbidities. It is conceivable that men with lower socio-economic status are less educated, seek medical advice at a later stage of the disease, and due to smoking have co-morbidities (diabetes, vascular diseases) that may hinder effective (surgical) treatment for cancer.

The association between obesity (elevated body mass index) and PC is also complex. A review article by MacInnis and English concluded that obesity is weakly associated with an increased risk for PC [MacInnis and English 2006]. The association is particularly strong in high-grade [Gong et al. 2006] and/or advanced stage PCs [Wright et al. 2007]. The increased risk could be attributable to problems with detecting PCs in obese men [Buschemeyer and Freedland 2007].

Of the dietary factors proposed to increase the risk for PC, there is some evidence for high processed meat consumption [Kolonel 2001], high zinc intake (as a supplement) [Leitzmann et al. 2003], folate and vitamin B12 [Collin et al. 2010] and high animal fat intake [Ma and Chapman 2009], but contradictory or insufficient evidence for dairy products, calcium and β-carotene [Ma and Chapman 2009]. Of the factors hypothesized to protect against PC, lycopene (a carotenoid found e.g. in tomatoes) has been identified as somewhat protective; whereas for soy (phyto-estrogens) there is less evidence [Wolk 2005; Ma and Chapman 2009]. Selenium and vitamin E have been proposed to protect against PC [Ma and Chapman 2009], but a large intervention trial (SELECT) found an increased risk for PC in men with supplementary vitamin E, but no change in risk for men with supplementary selenium or men taking both selenium and vitamin E [Klein et al. 2011]. So far there is no evidence for the protective effects of vitamin D [Gupta et al. 2009], vitamin C [Wolk 2005], fish [Terry et al. 2003] green tea, folate or omega-3 fatty acids [Ma and Chapman 2009].

Since inflammation may play a part in the development of PC, the effect of sexually transmitted diseases has also been evaluated. History of gonorrhea has been associated with a small increase in the risk for PC in a meta-analysis (OR

1.3, 95% CI 1.1-1.6), likewise syphilis (OR 2.3, 95% CI 1.3-3.9) [Dennis et al. 2002], but a recent prospective study from the Prostate, Lung, Colorectal and Ovarian Cancer screening trial (PLCO) found no association between sexually transmitted diseases and PC [Huang et al. 2008]; neither did another prospective study (the Health Professionals Follow-up Study) [Sutcliffe et al. 2006]. The evidence for the hypothesis that sexually transmitted diseases increase the risk for PC is, at best, inconclusive. A 30% reduced risk for PC has been identified in HIV-infected men, but this may be attributable to differential PSA screening [Shiels et al. 2011]. Finally, in several studies vasectomy has not been associated with increased risk for future PC [Lesko et al. 1999; Stanford et al. 1999].

In conclusion, the evidence for environmental risk factors for PC is scarce, although well-designed prospective studies and especially intervention studies would be welcome to better evaluate the role of the environment.

2.3.2 Primary prevention of prostate cancer

Given that PC is a substantial public health problem (see Section 2.4), major efforts have been made to find how PC could be prevented. Primary prevention means preventing the disease from occurring altogether, which is usually the most cost-efficient and also ethically sound method, when possible.

Drugs that inhibit 5α-reductase (the enzyme that converts testosterone into more potent dihydrotestosterone) are of special interest in preventing PC, since dihydrotestosterone is the hormone that stimulates the growth in both normal prostatic cells and cancer cells. There are two drugs that inhibit 5α-reductase, finasteride (inhibits type 2 isoform of the enzyme) and dutasteride (inhibits both type 1 and type 2 isoforms).

The Prostate Cancer Prevention Trial (PCPT) was a multicenter trial in the USA and Canada, and consisted of 18,882 men enrolled 1993-1997. The men were randomized to receive either finasteride or placebo for seven years. The trial was discontinued in 2003 for ethical reasons. Although overall PC risk was decreased in the finasteride group (18.4 vs. 24.4%), there was an increase in

high-grade cancers (6.4 vs. 5.1%; RR 1.27, 95% CI 1.1-1.5) [Thompson et al. 2003].

Retrospectively, the ramifications of the PCPT trial have been discussed in the scientific community. Three possible explanations have been proposed: 1) finasteride inhibits the development of low-grade cancers but stimulates the growth of high-risk cancers, 2) finasteride increases the detectability of high-grade cancers by decreasing the prostatic volume and increasing sensitivity and 3) finasteride alters the morphology of PC in a way that makes low-grade cancers appear high-grade to the evaluating pathologist [Lucia et al. 2007]. The second explanation has generally been accepted, and several analyses adjusting for this bias have been conducted. In all four studies the RR suggests no elevated risk for high-grade PC [Cohen et al. 2007; Pinsky et al. 2008; Redman et al. 2008; Kaplan et al. 2009]. Finasteride is no longer regarded as a drug that increases the risk for high-grade PC [Strope and Andriole 2010].

Dutasteride, another inhibitor of 5α-reductase, has also been investigated as a drug that may prevent PC [Andriole et al. 2010]. The Reduction by Dutasteride of Prostate Cancer Events (REDUCE) trial enrolled 8,231 men in 2003 mainly in Europe and North America. Two groups were formed: one to receive dutasteride daily and one to receive placebo. The selected men having PSA level 2.5–10 ng/ml were biopsied once within 6 months before enrollment and were free from PC. Of the men in the dutasteride group 19.9% were diagnosed with PC vs. 25.1% in the placebo group; relative risk reduction 22.8% (95% CI 15.2-29.8%) [Andriole et al. 2010]. Unfortunately, a significant risk reduction was only observed in Gleason score 5-6 cancers (low-grade), not in clinically more significant high-grade cancers – possibly due to the short follow-up time (4 years).

Statins, i.e. drugs that block 3-hydroxy-3-methylglutaryl coenzyme A in the cholesterol synthesis pathway, have also been proposed to protect against PC. The evidence is far from conclusive, although in a meta-analysis a protective effect against the risk for advanced PC was found (hazard ratio (HR) = 0.77; 95% CI 0.64-0.93) [Bonovas et al. 2008]. Based on the level of evidence so far, statins cannot be regarded as a preventive drug against PC [Rittmaster et al. 2009; Stephenson et al. 2010].

The effect of non-steroidal anti-inflammatory drugs on reducing PC risk is possible, but still inconclusive. Two meta-analyses have reviewed published data on the use of acetylsalicylic acid: one of these found a borderline statistically significant effect: HR = 0.89, 95% CI 0.77–1.03 [Jafari et al. 2009] and the other found a statistically significant effect: HR = 0.83, 95% CI 0.76–0.91 [Mahmud et al. 2010]. An analysis covering eight randomized trials found no statistically significant reduction of risk for PC death after at least four years of acetylsalicylic acid use (HR after >5 years of follow-up 0.52, 95% CI 0.20 – 1.34) [Rothwell et al. 2011]. Studies on the use of non-aspirin anti-inflammatory drugs (e.g. ibuprofen) have not shown a protective effect [Mahmud et al. 2010; Murad et al. 2011]. Given the many (potentially lethal) side-effects with long-term use of non-steroidal anti-inflammatory drugs their use to prevent PC cannot be justified, at least until more conclusive evidence from intervention studies is available.

At the moment the only sound evidence for the prevention of PC is for 5α-reductase inhibitors in preventing low-grade PCs [Stephenson et al. 2010; Strope and Andriole 2010]. Whether or not these drugs are cost-effective, increase quality-adjusted life years or reduce cancer-specific mortality is unknown. Although 5α-reductase inhibitors have not been proven to reduce the risk of lethal PC, preventing low-grade cancers may still be clinically very prudent, not only to avoid the adverse effects of overtreatment – and resulting costs – but to prevent anxiety related to cancer diagnoses.

2.3.3 Diagnosis and classification

2.3.3.1 Clinical diagnosis

The diagnosis of PC is always based on a histological sample from the prostatic tissue examined by a pathologist. Extracting this tissue is performed by a urologist when PC is suspected. Needle biopsies are generally taken from the prostate. Sometimes cancer cells are found in the extracted tissue material after

a transurethral prostatic operation for BPH and sometimes the diagnosis of PC is set in an autopsy.

Often there are no specific symptoms in PC. Moreover, lower urinary tract symptoms indicating an abnormal mass in the prostate may be due to a more common condition, BPH. Although PC may originate anywhere in the prostatic tissue, carcinomas are more common in the peripheral zone; BPH usually afflicts the central zone circulating the proximal urethra, which causes lower urinary tract symptoms. Since the conditions are not mutually exclusive but may both be present in the same prostate (and often are), careful diagnostics are needed to evaluate whether the patient has only BPH or also PC.

When PC has reached an advanced stage, systemic symptoms may be present: metastases to the bone may cause pain, there may be anemia or weight loss.

The most common basic examination is digital rectal examination (DRE) in which the physician palpates the prostate through the anus with a gloved finger. This is a readily available, inexpensive method of examining the status of the prostate. A suspicious DRE in most cases warrants biopsy. Factors arousing suspicion may be induration, a nodule or unilateral enlargement. A shortcoming of DRE is its low sensitivity, since potentially lethal cancers may be present but impalpable – in fact, the overall sensitivity of DRE has been estimated at 37% and specificity at 91% [Schröder et al. 1998]. The sensitivity, however, increases with increasing PSA level, since at PSA levels <3.0 ng/ml it is 20% and at PSA levels ≥3.0 ng/ml it is 46% [Schröder et al. 1998]. Moreover, DRE is dependent on the skill of the physician, and as a test it shows quite large interobserver variability [Gosselaar et al. 2008]. Given these findings, DRE still has its place in routine urologic work-up, but does not alone suffice to differentiate between benign and malignant disease.

Transrectal ultrasound (TRUS) is nowadays routinely performed when assessing the status of the prostate. With an ultrasound probe a urologist can visualize the prostate through the rectum wall, estimate the size of the prostate, assess the homogeneity of the parenchyma, and observe possible cancer growth through the capsule of the prostate. The benefits of TRUS are minimal invasiveness, relatively low cost and the fact that no ionizing radiation is

needed. The sensitivity of conventional TRUS has been estimated at 39-75% and specificity at 40-82% [Heijmink et al. 2011]. These wide ranges reflect the interobserver variability that is also a problem with TRUS. Using Doppler TRUS or contrast-improving agents improves sensitivity and specificity, but adds to the costs and makes the procedure more time-consuming [Trabulsi et al. 2010; Heijmink et al. 2011]. Conventional computerized tomography has currently no role in the diagnostics of PC due to the low resolution in soft tissues. Positron emission tomography has poor availability given the large number of PC candidates – and is moreover expensive. Magnetic resonance imaging (especially contrast-enhanced) may play a role in men with continuous suspicion of PC and several negative TRUS-guided biopsies [Jager et al. 2000; Heijmink et al. 2011].

The most common method of extracting histological samples from prostate is TRUS-guided needle biopsies. The procedure is fast, inexpensive and relatively easily available in an outpatient setting. The procedure however, is invasive and often causes the patient discomfort and moderate pain despite the use of local anesthetic. Minor adverse effects after a biopsy are common (84%), including hematuria, hematochezia and hematospermia, but the procedure is usually well-tolerated [Mäkinen et al. 2002; Carlsson et al. 2011]. A possible major adverse effect after prostate biopsy is infection (at best a simple urinary tract infection; at worst septicemia). Antibiotic prophylaxis has been deemed effective and necessary to reduce post-biopsy infections [Zani et al. 2011].

Originally sextant biopsy procedure was thought to be of sufficient sensitivity to diagnose PC [Hodge et al. 1989]. Since then, the observation that sextant biopsies may miss up to 30% of cancers [Norberg et al. 1997] has led authorities to cease to recommend sextant biopsies [European Association of Urology Guidelines on Prostate Cancer 2011]. Instead, taking at least 8-12 biopsies as posterior and lateral as possible is generally recommended. Increasing the number of cores to more than 12 does not appear to achieve additional sensitivity [Eichler et al. 2006]. However, using more cores (18-24) may be prudent if several negative biopsies have already been taken but a suspicion of cancer persists, or if the prostate in question is very large [Scattoni et al. 2010].

2.3.3.2 Biomarkers: Prostate-specific antigen and other tests

Prostate-specific antigen (PSA) is a serine protease (also known as kallikrein-related peptidase 3), excreted almost exclusively by prostatic cells. Its function is to liquefy the ejaculate so that the sperm can swim freely. Low levels of PSA are normally found in blood, but several prostatic conditions (inflammation, BPH, trauma, cancer) increase the systemic levels of PSA due to the perturbations in the normal histological architecture. Therefore PSA is an organ-specific but not disease-specific serum marker.

PSA was first identified and characterized in the 1960s and 1970s as a potential piece of forensic evidence for detecting seminal fluid [Hara et al. 1971]. Later it was found to be present in prostatic tissue and blood, and subsequently proposed as a marker for PC [Wang et al. 1981]. Prostatic cancer cells were found to leak PSA into blood much more than benign cells, and larger cancers leaked more PSA than smaller cancers [Stamey et al. 1987]. PSA quickly became of interest not merely as a tool for estimating cancer progression but as a diagnostic tool for prostate cancer. Subsequently it has been observed that a single PSA measurement at age 44-50 years predicts the risk for PC up to 25 years afterwards [Lilja et al. 2007].

The sensitivity and specificity of PSA as a diagnostic tool depend on the threshold. Even today there are only arbitrary limits as to what is the “normal” and what is an “abnormal” level of PSA, since the concentration of PSA is a continuous, not a dichotomous variable. A prospective study in 1991 used a cutoff point of 4.0 ng/ml, recommended by the manufacturer of the PSA assay (sensitivity was 79%, specificity 59%) [Catalona et al. 1991], and this limit has been used in many subsequent studies. The positive predictive value of PSA concentration above 4.0 ng/ml for cancer has been estimated at approximately 30% [Catalona et al. 1994; Määttänen et al. 1999], which conversely means that 70% of these men undergo unnecessary biopsies.

This cutoff point of 4.0 ng/ml, however, may leave as many as 15% of cancers undetected, as shown by an analysis from the PCPT trial. What is more, 15% of these cancers were Gleason 7 or higher (2.3% of all cancers) [Thompson et al. 2004]. Lowering the threshold for biopsy would improve sensitivity, but the

price for this is decreased specificity. From the PCPT trial data, it can be seen that for a 4.1 ng/ml threshold, sensitivity was 21% and specificity 94%, but for a 2.1 ng/ml threshold 53% and 73%, respectively [Thompson et al. 2005].

Instead of a single fixed cutoff point, age-adjusted cutoff points have been proposed to better tackle the tendency of PSA to increase with age (due to benign prostatic disorders). For example, Oesterling et al. proposed the following age-specific reference values: 40-49 years, 0-2.5 ng/ml; 50-59 years, 0-3.5 ng/ml; 60-69 years, 0-4.5 ng/ml; 70-79 years, 0-6.5 ng/ml [Oesterling et al. 1993]. Similarly, different reference values may be used for different ethnic groups (e.g. higher upper limits for African American men) [Greene et al. 2009].

Another method of improving the sensitivity and specificity of PSA is to use free/total PSA ratio (F/T PSA). Circulating PSA can be either free or in a complex between PSA and the protease inhibitor α1-antichymotrypsin. The proportion of free PSA has been observed to be smaller in men with PC than in men with a benign prostatic condition [Stenman et al. 1991], which has led to the notion that F/T PSA may be used as a tool in differentiating those with a malignant prostatic condition. Indeed, a review study in 2005 analyzed 66 studies and reported that the use of F/T PSA is particularly useful in PSA range 4-10 ng/ml, less so in the PSA range 2-4 ng/ml [Roddam et al. 2005]. Since then, in a population-based screening setting, the usefulness of F/T PSA in predicting future PC after an initial PSA level of <3.0 ng/ml has been validated [Finne et al. 2008]. Both F/T PSA (usually with a cutoff point of 16-18%) and age-specific reference values are commonly used in routine PC diagnostics in Finland.

PSA velocity (or PSA doubling time) has also been proposed to improve diagnostic accuracy. The rationale is to be concerned for patients in whom PSA increases over time, suggesting a malignant process. Although this seems a logical way to improve diagnostics, PSA velocity does not seem to substantially improve the prognostic value compared to a single PSA measurement [Ulmert et al. 2009; Vickers et al. 2009; Bangma et al. 2010]. PSA density (i.e. the ratio of total PSA to the volume of the prostatic gland measured by TRUS) is based on the notion that prostatic carcinoma tissue leaks more PSA to the bloodstream than does benign hyperplastic tissue. In theory, PSA density does improve the predictive values compared to PSA alone [Rommel et al. 1994], but the method

suffers from interobserver variability in estimating the volume by TRUS, which hampers its diagnostic use [Ulmert et al. 2009; Bangma et al. 2010].

Kallikrein-related peptidase 2, or hK2, is genetically and structurally close to PSA, and may be better than PSA in predicting organ-confined disease from non-organ confined disease, especially in PSA range <10 ng/ml [Raaijmakers et al. 2007]. A multivariable model utilizing total PSA, free PSA, intact PSA and hK2 improved the specificity for PC compared to just using total PSA and helped to reduce unnecessary biopsies by 57%, albeit also resulting in missing 20% of low-grade cancers and 8% of high-grade cancers [Vickers et al. 2008]. Several subforms of PSA have been proposed to improve PC diagnostics including benign prostatic hyperplasia-associated PSA (BPSA), p2PSA, nicked PSA and proPSA, all of which still require further validation in prospective studies being applied in clinical routine diagnostics [Bangma et al. 2010].

Finally, messenger RNA of prostate cancer antigen 3 (PCA3) is overexpressed in prostatic cancer cells, and has shown potential as a diagnostic tool for PC, particularly as an auxiliary test to improve specificity after the PSA test. As an independent test (i.e. replacing PSA) it is not sensitive enough [Roobol 2011]. PCA3 is measured from the urine after prostatic massage with DRE, and is not influenced by the size of the prostate (in contrast to PSA).

2.3.3.3 Grading and staging of prostate cancer

Cancers need to be classified in order to differentiate high-risk cancers from low-risk cancers to better guide treatment options and estimate prognosis. It is essential to ensure that low-risk cancer patients do not receive excessively aggressive treatment (in which case the treatment may cause more harm than the disease itself) and that high-risk cancer patients receive as aggressive treatment as is necessary. This is particularly important with PC, since two PCs can be extremely different diseases. It is customary to classify cancers by grade (how aggressive the cancer is histopathologically) and stage (how far the disease has spread).

PCs can be graded using several classifications, but the Gleason grading system has become predominant after being recommended by the World Health

Organization (WHO) in 1993 [Montironi et al. 2005]. Based on glandular architecture, the Gleason grading system produces a number corresponding to the most prevalent and second most prevalent cancer type found in the sample. These two numbers range from 1 to 5, and the sum of these two grades produces the Gleason score, ranging from 2 to 10. The higher the number, the less differentiated and thus the more aggressive the carcinoma is. Nowadays reporting a possible tertiary Gleason grade is recommended in radical prostatectomy samples, and in needle samples if the tertiary grade is higher than the primary and secondary. Gleason score is a well-established predictor of outcome in PC patients [Andrén et al. 2006]. In addition to Gleason scores, it is customary to use the classic WHO cancer grading system, which classifies cancers to grades I – III on the basis of nuclear anaplasia [Mostofi et al. 1980].

A familiar problem with the Gleason score is its tendency to lead to undergrading of PCs, i.e. low-grade cancers in biopsies are in fact of higher risk when re-examined after a radical prostatectomy [Egevad et al. 2001; Montironi et al. 2005; Iczkowski and Lucia 2011]. In addition, the agreement between pathologists is far from perfect: estimates range from 60-70% [Iczkowski and Lucia 2011].

There has been a substantial grade shift in Gleason scores in the recent years as new recommendations (Consensus Conference of the International Society of Urological Pathology in 2005) have been issued. For example, it is no longer recommended that Gleason score 2-4 be assigned in needle biopsies; and the criteria for cribriform pattern grade 3 have been tightened [Epstein et al. 2010]. These changes have made the previous Gleason score 2-5 cancers virtually nonexistent, as these cancers are more commonly graded Gleason score 6 or 7. For example, in a study where over 3,000 radical prostatectomy samples were reviewed applying the new recommendations, the proportion of Gleason score 6 cancers decreased from 48% to 22% and the proportion of Gleason score 7 cancers increased from 26% to 68% [Helpap and Egevad 2006]. Moreover, the concordance between biopsy specimen score and prostatectomy specimen score was improved from 58% to 72%.

Since grades 1-2 are not recommended and grade 5 (the least differentiated) is fairly rare, grades 3 and 4 (basically Gleason scores 6 - 8) form the prognostic

watershed area of PC. The difference between Gleason score 3 + 4 = 7 and 4 + 3 = 7 is especially marked: the latter had a 3.1 times greater risk of death (95% CI 1.1-8.6) compared to the first during 20 years of follow-up [Stark et al. 2009].

PC staging is based on the classic TNM staging system originating in the 1940s and 1950s. A TNM stage consists of three components: T = the extent of the primary tumor, N = the absence or presence of lymph node metastases and M = the absence or presence of distant metastases (Table 2.1) [International Union Against Cancer 2009]. Generally, T1-2N0M0 cancers can be considered local, T3-4N0M0 locally advanced and T1-4N0-1M1 advanced.

TNM stage is clinically estimated at the time of diagnosis (cTNM), whereas a pathological (pTNM) can be reached only after surgical treatment (in which actual tissue samples are obtained). Estimating the cTNM is important because it affects the available treatment modalities and prognosis; it is particularly important to evaluate whether the cancer is intracapsular (T1-2), extracapsular (T3-4) or has metastasized (M1).

cT is based on DRE and TRUS; a T1 cancer is impalpable and macroscopically invisible, and is found by chance (prostate resection for BPH) or an abnormal PSA test. Larger tumors can be felt in DRE or seen in TRUS. However, the important differentiation between T2 and T3 disease is often difficult. Serum PSA level or its derivatives cannot differentiate between the two with sufficient accuracy [European Association of Urology Guidelines on Prostate cancer 2011]. Magnetic resonance imaging (especially endorectally), positron emission tomography and novel TRUS techniques (contrast medium, Doppler) have not gained a foothold in routine PC staging.

The lymph node status can only be assessed after surgical lymphadenectomy – imaging studies are not sufficiently accurate, as there may be small metastases even though the nodes are not enlarged [Heesakkers et al. 2008]. Patients with cT≤2, PSA <20ng/ml and Gleason score ≤6 have a less than 10% chance of having lymph node metastases [Partin et al. 2001].

Since PC commonly metastasizes to the bone, a bone scan (scintigraphy) is often performed to assess the existence of possible distant metastases. In Finland it is recommended to perform a bone scan in patients with either

Table 2.1. TNM classification of prostate cancer [International Union Against Cancer, 2009].

T	**The extent of the primary tumor**		
	Tx		The primary tumor cannot be assessed
	T0		No evidence of primary tumor
	T1		Clinically inapparent tumor, neither palpable or 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, e.g., because of elevated prostate-specific antigen (PSA)
	T2		Tumor confined within the 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 prostatic capsule
		T3a	Extracapsular extension (unilateral or bilateral) including microscopic bladder neck involvement
		T3b	Tumor invades seminal vesicle(s)
	T4		Tumor is fixed or invades adjacent structures orher than seminal vesicles: external sphincter,
			rectum, levator muscles and/or pelvic wall
N	**Regional lymph nodes**		
	Nx		Regional lymph nodes cannot be assessed
	N0		No regional lymph node metastasis
	N1		Regional lymph node metastasis
M	**Distant metastasis**		
	M0		No distant metastasis
	M1		Distant metastasis
		M1a	Non-regional lymph node(s)
		M1b	Bone(s)
		M1c	Other site(s)

Gleason 8-10 cancer, PSA above 20 ng/ml, elevated serum alkaline phosphatase or if the patient has skeletal pain suggesting bone metastases [Käypä hoito: Eturauhassyöpä 2007].

2.3.4 Treatment and adverse effects

Due to the wide clinical spectrum of PCs, a wide spectrum of treatment options is available. In high-risk cancers, the treatment schemes are relatively straightforward, and the choice of treatment can be made in concordance with the patient, weighing the benefits and downsides of the treatment modality. A PC is curable when it is still confined inside the prostatic capsule. There is clear evidence that high-grade PCs (even those that are clinically localized at diagnosis) are lethal if not treated aggressively: in 20 years of follow-up, the mortality rate was 30/1000 person-years for Gleason 6 cancers, 65/1000 for Gleason 7 cancers and 121/1000 person-years for Gleason 8-10 PCs [Albertsen 2005]. Furthermore, these results were influenced by the grade shift, i.e. many of the Gleason 6 cancers would nowadays be graded Gleason 7.

The present-day challenge both in screening and in clinical work arises from the low-risk cancers: small, indolent, low-grade cancers that are commonly found in wild screening with PSA. All PCs eventually progress, but some progress so slowly that they never actually affect morbidity or mortality. Treating all PCs aggressively would result in massive overtreatment, and hence focusing aggressive treatment for men with intermediate to high- risk cancer is not only medically but also ethically prudent.

PC treatment guidelines are readily available both on national level in Finland [Käypä hoito: Eturauhassyöpä 2007] and on international level in Europe [European Association of Urology Guidelines on Prostate Cancer 2011].

The various treatment options and their adverse effects will be outlined only briefly in this section.

With watchful waiting, no active measures are taken to cure the cancer or to impede its progression unless the patient develops symptoms that may justify surgical intervention (such as transurethral resection of the prostate to alleviate

urinary retention, or palliative radiation treatment for painful bone metastases). Watchful waiting may be chosen if the patient is elderly, has a low-risk PC or his general condition is too weak to allow more active measures.

Active surveillance is a more modern approach, and also a conservative option for low-risk cancers. With active surveillance, possible cancer progression is monitored actively by PSA tests and rebiopsies – the purpose is to differentiate which cancers are progressive and to catch these in time for curative treatment. This method is also appropriate for younger, relatively fit men who could undergo radical treatment but in whom the disease progression risk is so low that radical treatment might cause more harm than the disease itself. An adverse effect with active surveillance is anxiety related to waiting whether the cancer progresses or not, which may become overwhelming for some, who then choose aggressive treatment. Structured schemes for active surveillance help to control the anxiety of both the patient and the physician, and studies on the effectiveness of active surveillance are ongoing [Cooperberg et al. 2011]. Long-term results are not yet available on the effectiveness of active surveillance, but there is preliminary evidence that at median follow-up of less than 10 years active surveillance schemes produce disease-specific mortality at 1-2.5% for low risk PCs and 4-5% for intermediate risk PCs [Stattin et al. 2010; Bul et al. 2012].

Curative treatment options include radical prostatectomy, external radiation treatment and brachytherapy. Radical prostatectomy means surgically removing the entire prostatic gland and seminal vesicles, often with a bilateral pelvic lymph node dissection, and constructing an anastomosis between the urethra and the bladder neck. The purpose is to remove all the cancer (i.e. reach negative surgical margins), while preserving continence and, when possible, potency. Robotic-assisted radical prostatectomies are increasingly often performed in Finland, but have not yet yielded better long-term results. Radical prostatectomy can be performed if the cancer has not spread through the capsule of the prostate (T≤3a), i.e. curative treatment is still possible, and usually if the patient has more than 10 years of expected lifetime. In addition to any adverse effects related to major surgery (perioperative death, major bleeding, infections), the adverse effects of radical prostatectomy include

incontinence (depending on definition in 5-10% of patients) and impotence (10-60%) [Käypä hoito: Eturauhassyöpä 2007; Ficarra et al. 2012].

Radiation therapy is another option for curative treatment, and can be administered either externally (external beam radiation therapy) or internally (brachytherapy). In brachytherapy, radioactive seeds are inserted under ultrasound-guidance into the prostate. Immediate adverse effects after radiation therapy include nocturia, incontinence, dysuria, frequent need to urinate and additionally symptoms from the gastrointestinal tract (diarrhea, frequent need to defecate) may be present. These adverse effects usually dissipate in 2-3 months, but permanent adverse effects may occur later within months or years after therapy: nocturia, urge incontinence, rectal bleeding, diarrhea and impotence (in approximately 50% of men) [Käypä hoito: Eturauhassyöpä 2007].

Hormonal therapy is an option when the cancer cannot be curatively treated, i.e. when it has advanced locally or metastasized. This can also be used as a neoadjuvant or adjuvant therapy. Since the cancer cells depend on androgens, blocking them usually halts the growth of the tumor. Bilateral orchiectomy (removal of the testes) stops the production of testosterone (and hence, the more potent dihydrotestosterone). A similar effect can be achieved using drugs that inhibit the production of testosterone or by using anti-androgen drugs which inhibit the stimulating effect of circulating androgens. Depriving the androgens often results in symptoms such as hot flushes, nocturnal sweating, impotence and loss of libido. These may have a decidedly deleterious effect on quality of life.

So far there is no evidence from randomized controlled trials on the efficacy and long-term results between radical prostatectomy, external beam radiation therapy or brachytherapy, and the choice between the options must be made in concordance with the patient weighing up the benefits and downsides [European Association of Urology Guidelines on Prostate cancer 2011].

There is evidence for the superiority of radical treatment vs. watchful waiting (in low-risk T1-2 cancers) from a randomized controlled trial, Scandinavian Prostate Cancer Group Study Number 4, in which both cancer-specific mortality (RR 0.62, 95% CI 0.44-0.87) and all-cause mortality (RR 0.75, 95% CI 0.61-0.92) were lower in men randomized to radical prostatectomy at 12.8 years of follow-

up [Bill-Axelsson et al. 2011]. However, this trial was conducted mostly before wild PSA testing became common and may not be applicable to the present day.

2.4 Occurrence of prostate cancer

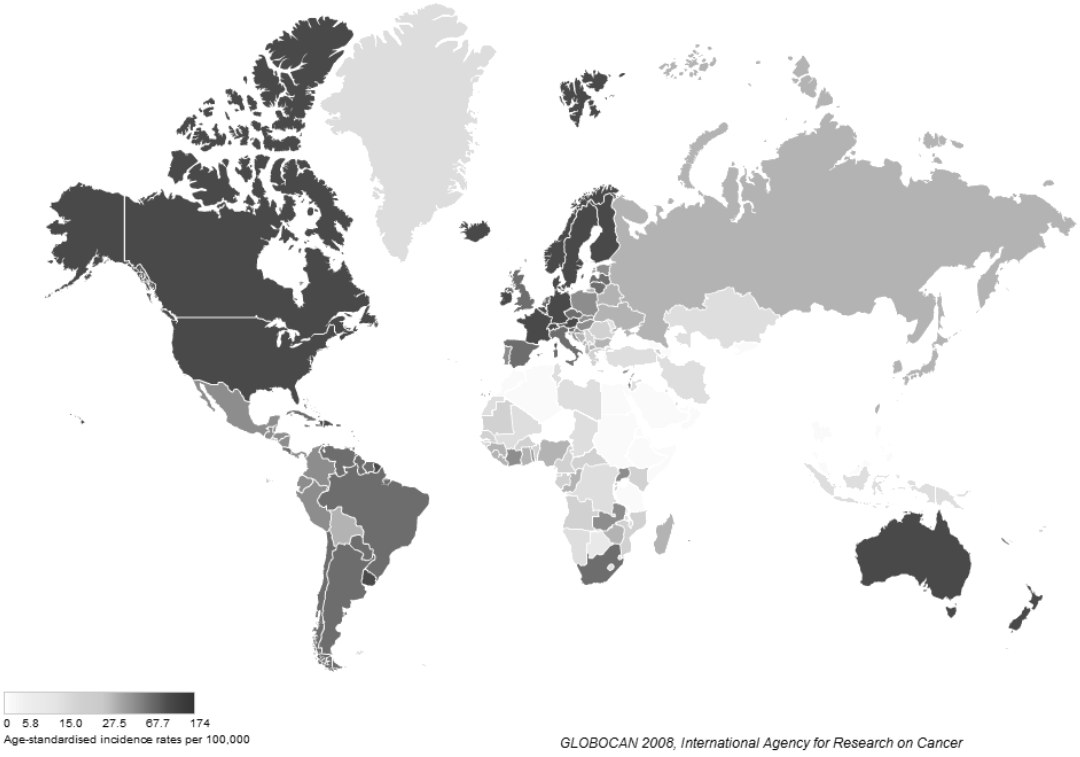
2.4.1 Incidence

Prostate cancer is currently the most common non-skin cancer in Europe, with an estimated 379,000 new cases in 2008 [Ferlay et al. 2008]. In Finland, 4,600 men are diagnosed with PC annually, and the cumulative risk of being diagnosed with PC before age 75 years is 11.4% [Engholm et al. 2011].

Roughly 70% of the registered PCs in the world are estimated to be diagnosed in the developed countries, and there are as large as 25-fold differences in the incidence of PC worldwide (Figure 2.3). The highest incidence rates (age-standardized rates per 100,000 in 2008) are seen in Northern Europe (Finland: 83.2; Sweden: 95.5), Australia (105.0) and the USA (83.8); whereas lowest rates are observed in Asia (China: 4.3; Vietnam 3.2; India 3.7) [Ferlay et al. 2008].

Several factors may explain why the incidence rates vary so widely internationally. Firstly, differences in the validity of data may explain the large differences. Comprehensive cancer registration requires substantial continuous effort from the authorities. In some countries (such as in Finland), cancer registration is nation-wide and population-based, all citizens have social security numbers, and it is obligatory for medical doctors to report new cancer cases. In Finland, 99% of solid tumors are reported to the Finnish Cancer Registry [Teppo et al. 1994], but due to incomplete data in many other countries cancer incidence is estimated by the incidence in a single city or when no data are available, by the rates in neighboring countries [Ferlay et al. 2008]. Only when cancer registry data are sufficiently valid is comparing international rates sensible. Since the incidence of PC varies so much even between highly

Figure 2.3. Age-standardized prostate cancer incidence rates / 100,000 varied considerably in 2008. [Ferlay et al. 2008]



developed countries (e.g. Japan and the USA), it seems reasonable to assume that the differences in incidence are to some extent real and not merely artifactual.

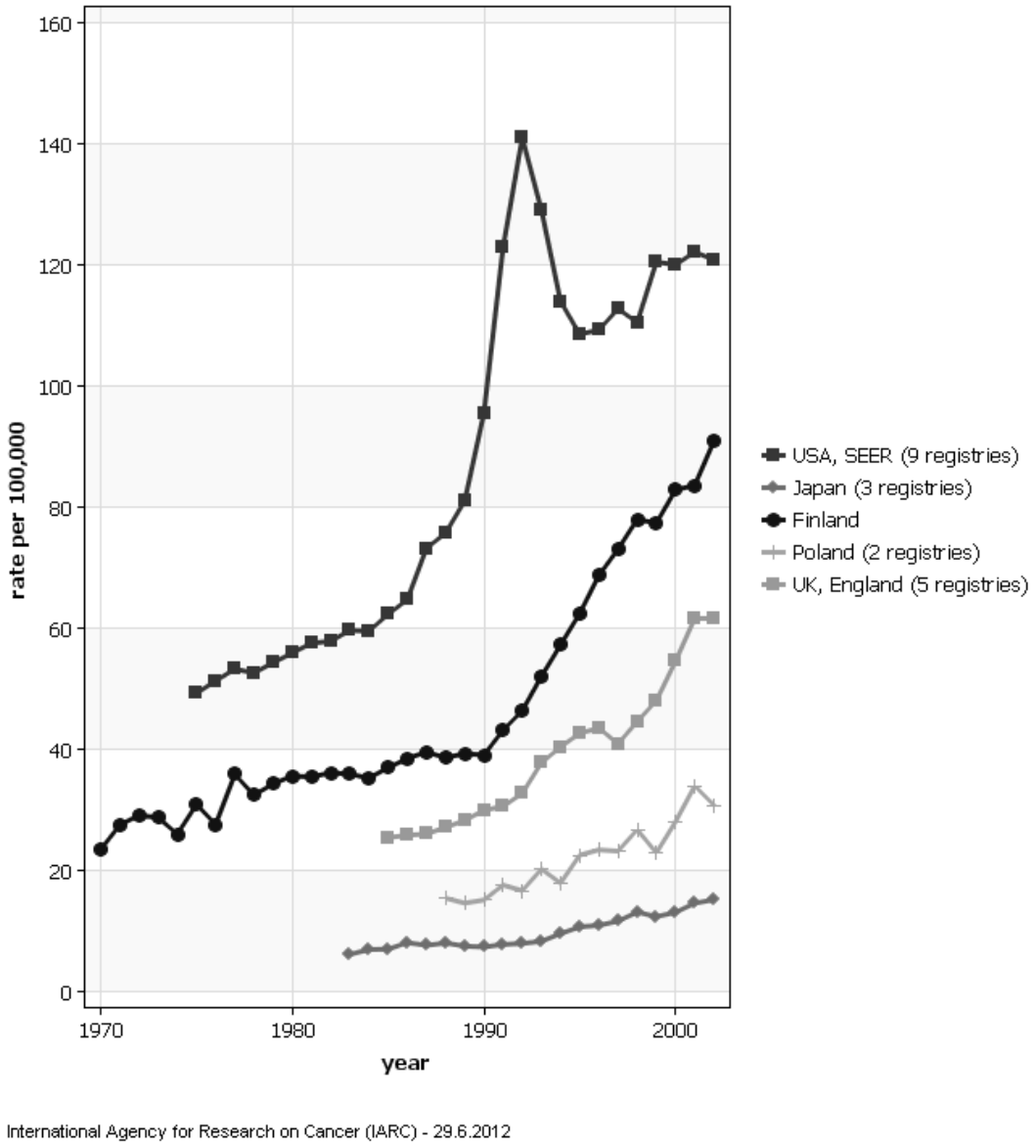
Secondly, genetic or environmental factors may explain why there is more PC in some countries than others, as reviewed in Section 2.3.1. Interestingly, despite the differences in the incidence of clinical PC, the prevalence of subclinical foci of latent PC seems to be similar around the globe [Breslow et al. 1977]. The relation of genetics and environment to the development from these latent foci to manifest disease is largely unknown. Thirdly, PC is very much an age-dependent disease, hence the aging population in developed countries (especially in Europe) could in part explain the high incidence rates. Global cancer rates are, however, age-standardized, i.e. age cannot explain the differences in cancer incidence, but an aging population naturally does contribute to the absolute number of PC cases.

Fourthly, societal, cultural, and behavioral differences (e.g. in seeking medical advice, the availability of medical services, differences in diagnostic options) affect PC incidence very much. This can be seen from the incidence trends from the 1970s to the early 21st century (Figure 2.4). For example, in the USA the annual percent change in PC incidence during the period of 1983-2000 was 4.7%, in Finland 5.7% and in Japan 4.3% [Curado et al. 2007]. This rapid increase in incidence seen in most developed countries cannot be explained by genetic factors or aging, and in this magnitude is not likely to be caused by environmental risk factors. Changes in cancer registration protocols could cause substantial shifts in incidence trends, but no such changes can account for the shift in PC incidence. The rapidly increasing phase of incidence begins in the late 1980s, which is when PSA testing for PC diagnostics became common.

Is the rapid increase attributable to PSA-testing? We cannot measure the extent of PSA testing retrospectively, but some estimates have been published. In the USA in 2001, 75% of men aged over 50 years had had a PSA test in their lifetime according to a questionnaire study [Sirovich et al. 2003]. By contrast in the Netherlands in the period of 1997-2000, roughly 20% of men in the control arm of a screening study were tested with PSA during that period [Otto et al. 2003]. Furthermore, in a questionnaire study 18% of Finnish physicians reported regular PSA screening for asymptomatic men in 1999 [Pogodin-Hannolainen et al. 2011]. Although these figures are not directly comparable, they reflect the fact that PSA testing is not equally widespread across all countries.

There is no doubt that PSA testing has had a major effect on the increase in PC incidence, but other factors, such as increased awareness of PC, increased number of transurethral resections of the prostate and improved diagnostic methods (TRUS-guided biopsies) have surely also contributed. In fact, a study in the Netherlands showed that the increase in PC incidence in the early 1990s was mostly due to T2 cancers (not PSA-detected, often T1c) and only the second peak in the early 21st century was due to asymptomatic, PSA-detected T1c cancers [Cremers et al. 2010]. An indisputable fact is, however, that PCs have undergone a clear stage shift, i.e. they are more likely to be localized at the time

Figure 2.4. Age-standardized prostate cancer incidence trends have been increasing worldwide (USA, Japan, Finland, Poland and UK) [Curado et al. 2007].



of diagnosis than e.g. in the 1980s, indicating earlier diagnosis [Miller et al. 2003].

In the very last years, PC incidence has been decreasing in many countries, including Finland, the USA, and the Netherlands [Ferlay et al. 2008]. It is not known conclusively why, but it is possible that PC incidence has reached saturation point, i.e. all the active measures of previous years have ultimately found most of pre-clinical phase PCs [Bray et al. 2010].

2.4.2 Mortality

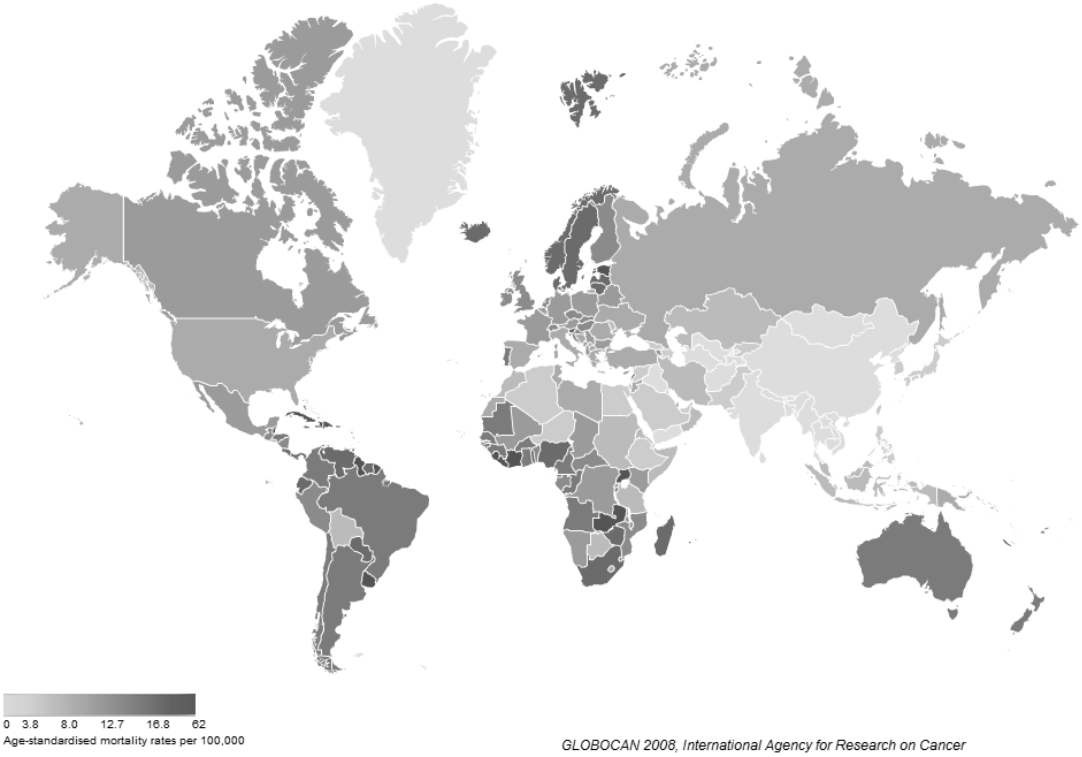
PC mortality also differs considerably between geographical areas, but not as dramatically as incidence. In Europe 94,000 men died of PC in 2008. In Finland, 800 men die annually of PC, which is 14% of all cancer deaths. The cumulative risk for dying of PC before age 75 years is 1.1% [Engholm et al. 2011].

Although in 2008 over 70% of PC diagnoses were made in developed countries, only 53% of PC deaths occurred in developed countries [Ferlay et al. 2008]. High mortality rates have been observed in the Scandinavian countries (Norway 18.6/100,000; Sweden 19.9; Denmark 19.8), Baltic countries (19.3-22.0/100,000), Caribbean countries (~25/100,000) and several African nations (Southern Africa 19.3/100,000) (Figure 2.5). Low mortality rates are again found in Asia: e.g. India 2.5; China 1.8; Republic of Korea 4.1 [Ferlay et al. 2008].

Thus, although there are 25-fold differences in incidence, only 10-fold differences in mortality from PC have been observed. Moreover, the incidence to mortality ratio in PC is roughly 8:1, whereas for lung cancer it is only 1.3:1 and for colorectal cancer 2.1:1 [Albertsen 2011]. This could suggest a substantial amount of overdiagnosis in several developed countries, meaning that many clinically insignificant PCs are diagnosed. The extent of overdiagnosis in PC has been estimated to be 23-42% of all PCs [Etzioni et al. 2002; Draisma et al. 2009]. A high incidence to mortality ratio could also indicate that treatment for the cancer is exceptionally efficient, but as the ratio has changed so fast it is unlikely to be attributable solely to better treatment modalities.

Mortality trends have also undergone changes in recent decades, as shown by Figure 2.6. In many developed countries the peak mortality rate was seen in the 1990s and thereafter a decline has been observed. The evidence for the effect of PSA on the decline in mortality is inconclusive [Weir et al. 2003; Etzioni et al. 2008; Bray et al. 2010]. For example, mortality has also declined in countries with relatively little PSA testing, and conversely, mortality has not declined in countries with widespread PSA testing [Oliver et al. 2001; Collin et al. 2008; Jemal et al. 2010]. More advanced treatment modalities have also been suggested to underlie the decline in mortality. While it is true that radical

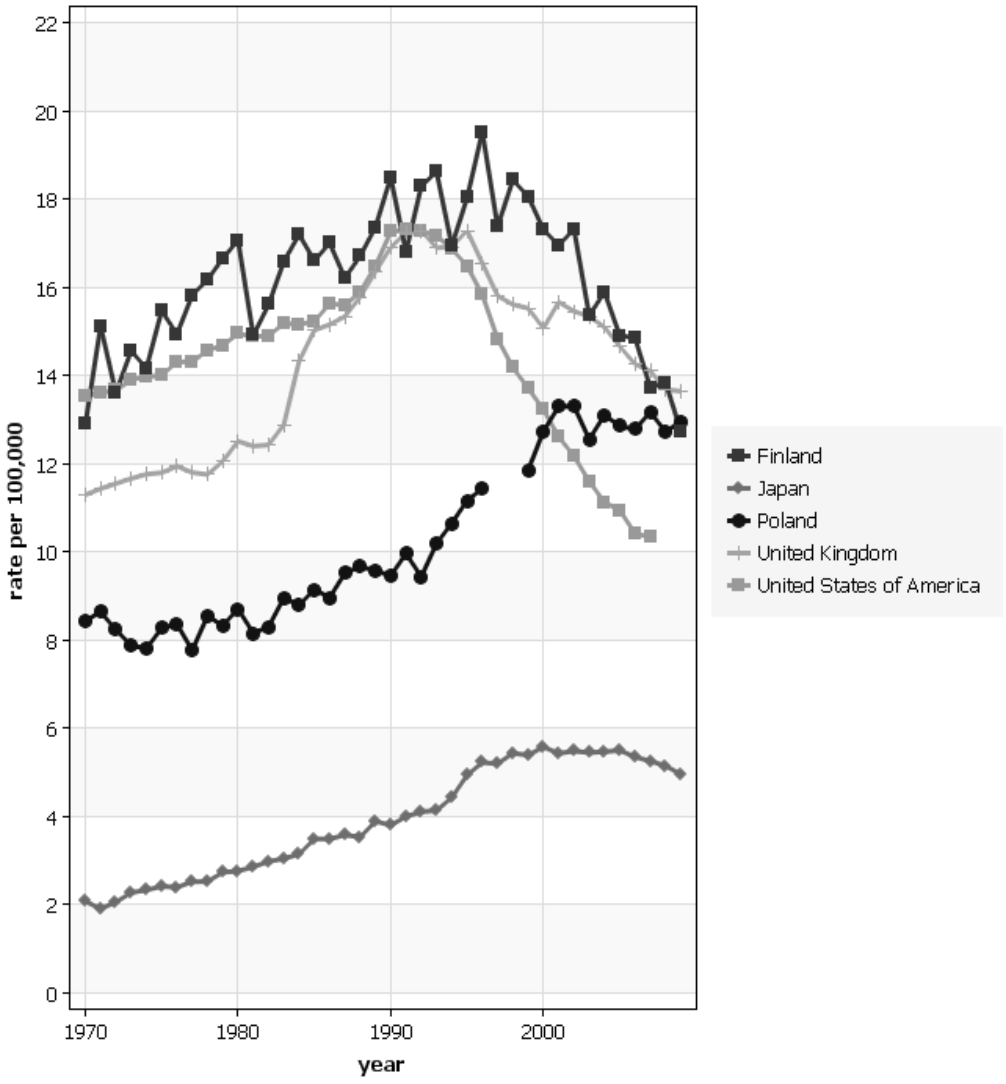
Figure 2.5. Age-standardized prostate cancer mortality rates / 100,000 globally in 2008. [Ferlay et al. 2008]



treatment is currently more common than before the PSA era (when a large part of PCs were beyond radical treatment), a simulation model showed that neither effective treatment nor PSA testing can be attributed solely to the decline in mortality [Etzioni et al. 1999].

It seems evident that the observed reduction in mortality is due to several factors: earlier detection due to PSA and surgical treatment for BPH, and more advanced treatment options.

Figure 2.6. Age-standardized prostate cancer mortality trends / 100,000, from 1970 to 2009 have been decreasing in several countries (Finland, Japan, Poland, United Kingdom, and USA) [World Health Organization 2011].



International Agency for Research on Cancer (IARC) - 29.6.2012

2.5 Cancer screening

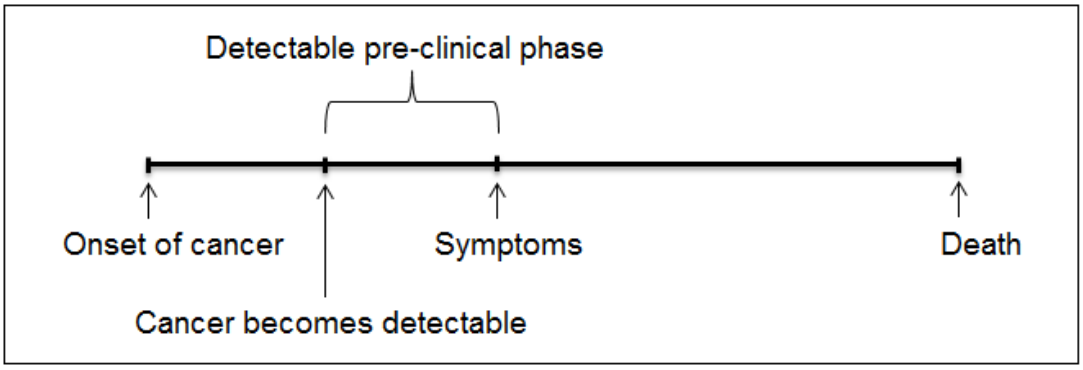
The purpose of cancer screening is to prevent deaths from cancer by reaching diagnosis at an early (asymptomatic) stage when the disease is still curable. Cancer screening may have other beneficial effects in addition to prolonging life, such as improving the quality of life by avoiding burdensome treatments, and lowered economic costs to society if the treatment of disease caught earlier is less expensive than that of an advanced disease. However, these benefits should be proven before systematic screening programs are initiated, because screening always carries negative consequences, as discussed later.

The rationale for screening is often based on the classic criteria set by a WHO workgroup in 1968 [Wilson and Jungner 1968]. These criteria can be summarized in six questions [quoted with permission, from Bhopal 2008]:

i. Is there an effective intervention?
ii. Does intervention earlier than usual improve outcome?
iii. Is there an effective screening test that recognizes disease earlier than usual?
iv. Is the test available, affordable, and acceptable to the target population?
v. Is the disease one that commands priority?
vi. Do the benefits of screening exceed the costs in this society?

The only way to answer these questions in a valid manner is through randomized controlled trials, and even after that, the fifth and sixth question must be assessed. It is not uncommon that large-scale medical interventions are launched without conclusive proof of their benefit-harm ratio or cost-effectiveness; instead, decisions may be made based merely on assumptions from clinical practice or insufficient scientific evidence. It is particularly difficult to evaluate the effectiveness of an intervention when the intervention (a screening test) has already spread to common clinical practice, since the established control group is easily subject to the same intervention outside the trial (contamination effect).

Figure 2.7. On a timeline from the onset of cancer to possible death, the window of opportunity in cancer screening is the detectable pre-clinical phase.



In a cancer screening program, asymptomatic people are tested with an appropriate screening test to find out who either have the disease (e.g. cervical cancer screening) or who are likely enough to have disease to justify further examinations (e.g. prostate, lung, colorectal and breast cancer screening).

The aim of screening is to detect cancer at a pre-clinical phase, i.e. after the onset of the disease but before clinical symptoms (when the person would have sought medical attention regardless of screening) (Figure 2.7). This is the only window of opportunity for screening programs, and the length of this detectable pre-clinical phase depends on the type and aggressiveness of the cancer and the sensitivity of the diagnostic test.

While the intentions of screening are good, screening always has negative consequences, possibly even more than positive consequences [Hakama and Auvinen 2008]. Firstly, screening tests are not perfect and may yield a false-negative or a false-positive screening result. A false negative result offers false reassurance of being free from disease, which may in turn delay correct diagnosis. An FP result not only causes undue anxiety about a disease that is nonexistent but may also subject these people to unnecessary invasive examinations that may have adverse effects. Secondly, cancer screening does not only identify cancers that are clinically significant and cause premature death. Screening often identifies a wide spectrum of lesions from premalignant states to indolent cancers to high-risk cancers. Not all cancers in e.g. the mammary glands and prostate affect the life expectancy or increase morbidity –

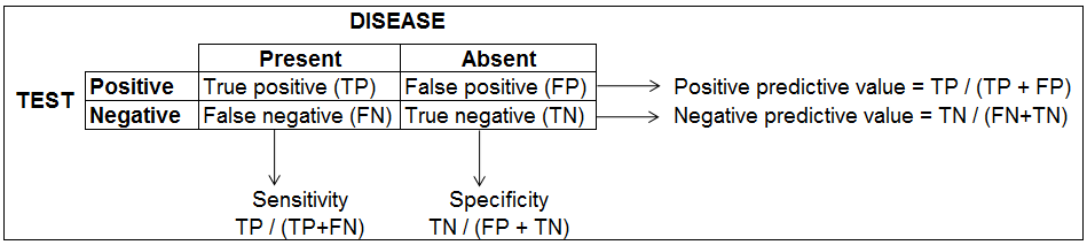
thus diagnosing these cancers and treating them aggressively results in overdiagnosis and overtreatment respectively. Another example of overdiagnosis is diagnosing cancers in subjects who will die of other causes before the cancer would have killed them – in which case the diagnosis was of no benefit, either. Finally, screening programs are expensive and require substantial amounts of resources over long periods. Whether or not these resources could be better allocated elsewhere demands careful consideration.

Screening requires particularly sound judgment in terms of harms and benefits, because screening predominantly affects healthy people. In a systematic screening program, the health authorities initiate the contact and invite people to participate – in contrast to clinical work, where the patient initiates the contact (and usually has sufficient symptoms and maladies to seek medical attention). The responsibility of afflicting adverse effects of screening on healthy people is thus particularly heavy [Bhopal 2008].

The validity of a screening test (or any diagnostic test) can be measured in terms of sensitivity and specificity (Figure 2.8). Sensitivity refers to the proportion of true positives of all those who actually have cancer, i.e. the ability of the test to identify those afflicted by cancer. With low sensitivity many cancers are missed. Sensitivity may differ depending on definition: *test sensitivity* refers to the ability of a single screening test, episode *sensitivity* refers to sensitivity corrected for non-attendance and selection, while *program sensitivity* takes into considerations not only the aforementioned factors, but also how the screening program succeeds in referring screen-positive cases to diagnostic examinations and in covering the intended population [Hakama et al. 2007]. From the same trial, test sensitivity may be 85%, episode sensitivity 48% (due to non-attendance) and program sensitivity 36% [Hakama et al. 2007].

Specificity, in turn, refers to the proportion of true negatives from all those who are free of cancer, i.e. the ability of the test to identify those free of cancer [Hakama and Auvinen 2008]. With low specificity the test is positive in many subjects who are in fact cancer-free. Sensitivity and specificity depend on the cutoff point of a continuous test variable (such as PSA or cholesterol level), and can be plotted on a receiver operating characteristic curve to optimize sensitivity and specificity. In other words, sensitivity and specificity are two

Figure 2.8. The relationship of sensitivity, specificity, positive predictive value, and negative predictive value can be presented in and calculated from a classic 2x2 table.



sides of the same coin: when the one increases, the other decreases (e.g. if a PSA level of 100 ng/ml is used as a cutoff point, the test is very specific but very insensitive, but the opposite is true if a PSA level of 0.5 ng/ml is used as a cutoff point).

Two other concepts are also used to assess the validity of a screening test: positive predictive value (PPV) and negative predictive value. The first refers to the proportion of true positive subjects of all those who have tested positive. In other words, it tells how likely a subject is to be really diseased if the test shows positive. Negative predictive value refers conversely to how likely a subject who has tested negative is really cancer-free. Predictive values are dependent on the prevalence of disease in the test population.

The selection of an appropriate screening test is a major decision in designing a screening program. On the one hand sensitivity, specificity, positive and negative predictive values should be maximized, but on the other hand a test that is inexpensive, noninvasive and not harmful should be chosen. For example, while flexible colonoscopy is relatively good at finding colorectal cancer, it is usually out of the question as a mass screening test due to its invasiveness and cost. Another example could be high-resolution computerized tomography, which is sensitive for lung cancer but uses a relatively high dose of radiation. The absolute extent of the adverse effects of such screening tests can be reduced by selecting a target population with a high *a priori* prevalence of lung cancer (e.g. middle-aged long-term tobacco smokers) [National Lung Cancer Screening Trial Research Team 2011].

The effectiveness of a screening protocol is not necessarily easy to measure. The main outcome in a cancer screening protocol is mortality, which can mean either overall mortality (from any cause) or cancer-specific mortality. While cancer-specific mortality is typically of interest in evaluating the efficiency of a cancer screening program, cancer-specific mortality may be subject to bias if the causes of death are misclassified. All-cause mortality, by contrast, concerns only the occurrence of death and is thus more reliable. It can be argued that preventing cancer-specific deaths is of no great importance if the subjects will die of other causes; and also that an observed reduction in cancer-specific mortality is not a strong piece of evidence if the overall mortality is not reduced as well [Black et al. 2002]. Since the accumulation of mortality data is slow (often more than 10 years), surrogate measures such as the incidence of metastasized cancer can be used to evaluate the effectiveness of screening, but these surrogate measures cannot provide conclusive evidence.

Cancer screening trials are also subject to three major biases. Firstly, a selection bias in the form of "healthy screenee bias" occurs when those who participate in screening are different (e.g. more health-conscious) from those who choose not to participate [Zeliadt et al. 2007]. This bias is particularly strong in volunteer-based trials. Although in properly executed volunteer-based trials the bias is not different between the study arms (subjects are randomized after consent), it makes it more difficult to generalize the results to general population (e.g. the participation rate is higher than if screening were to be implemented at national level). Only population-based randomized trials that analyze results on the intention-to-screen principle (i.e. all men are analyzed in the screening arm regardless of participation) are free of this healthy screenee bias.

Secondly, a lead time bias is often present in cancer screening trials [Hakama and Auvinen 2008]. Lead time refers to the time that the diagnosis is advanced by screening (in contrast to the time the diagnosis would have been made in a routine clinical setting without screening) [Hutchison and Shapiro 1968]. The maximum lead time is the length of the detectable pre-clinical phase. Lead time bias occurs when screening advances the date of diagnosis without actually postponing death, i.e. early diagnosis does not improve prognosis. However, due

to the earlier diagnosis it appears as if the time from diagnosis to death is longer than in non-screened cancers. This would falsely make cancer screening seem to postpone death (although it only increases survival time), and therefore survival time after diagnosis is not a suitable measure. Lead time bias can be estimated and controlled for with relatively complicated modeling [Draisma et al. 2009], but cannot be directly observed since we cannot know when the disease would have been diagnosed without screening.

Length bias is the third major bias, and potentially the most difficult to control for [Hakama and Auvinen 2008]. It results from the fact that screening is more likely to detect slow-growing cancers (which have a longer detectable pre-clinical phase) than aggressive cancers [Feinleib and Zelen 1969]. For example, if a cancer is screened for every five years, the aggressive cancers are more likely to surface clinically between screens than are slow-growing cancers. This artificially improves the survival in the screened cancer cases. Length bias can be avoided by including the entire randomization arm in analyses, including interval cancers detected outside the screening protocol.

2.5.1 Prostate cancer screening

PC screening meets many of the criteria set for mass screening (see section 2.5). There is an effective intervention and treatment for PC (i), and intervention earlier than usual improves outcome (ii). There is a relatively effective test that recognizes most cancers earlier than usual (PSA and F/T ratio) (iii), and this test is available, affordable and acceptable in most countries (iv). PC is a major public health challenge, as it is common and potentially lethal (v). The question that remains unanswered is whether the benefits of PC screening exceed the costs (vi), which can be understood as not only financial costs but also harm in the form of lost quality of life due to overdiagnosis and overtreatment.

There are some specific issues which make screening for PC with PSA particularly challenging. PCs tend to have long latent periods before surfacing clinically, if indeed they surface at all. Moreover, subclinical PCs are common (30-50%) in men aged >50 years and the incidence increases with age [Sakr et

al. 1994]. These observations, combined with the fact that PSA as a screening tool has better sensitivity than specificity (when sensitivity is measured by clinically relevant disease), result in a potentially behemothic amount of overdiagnosis resulting in overtreatment. With PSA, sensitivity and specificity in a screening setting are not easy to define, as the use of such concepts requires knowledge of which patients have cancer and which do not. Classifying men with low PSA concentration as "cancer-free" without biopsies is problematic, as these men can also have subclinical PCs (even clinically relevant ones, as shown by the PCPT trial) [Thompson et al. 2004]. Therefore estimating sensitivity and specificity for PSA as a diagnostic test would require biopsying all men regardless of PSA concentration. Although prostate biopsies are the gold standard for diagnosing PC, even they miss cancers.

Mass screening for PC with PSA has been under scrutiny for over 20 years, but nationwide screening has not been undertaken in any country. Many case-control studies and ecological studies have been published, yielding inconsistent results, but only randomized controlled trials can show conclusively whether screening for PC with PSA reduces mortality. The first randomized controlled trial was conducted in Québec, Canada, starting in 1988 and using a PSA cutoff 3.0 ng/ml (Table 2.2) [Labrie et al. 1999; Labrie et al. 2004]. This study suffered from a low participation proportion in the screening arm (only 24%), and an unknown level of contamination in the control arm (at least 7.3%). With a follow-up of over eight years, they showed that PC mortality was lower in the men who were screened, RR 0.39 (95% CI 0.19-0.65) [Labrie et al. 1999]. This figure is, however, calculated of the men who actually were screened (regardless of the randomization arm), and an intention-to-screen analysis yielded RR 1.01 (95% CI 0.76-1.33), i.e. no difference between study arms. Another randomized controlled (pilot) trial in Sweden was initiated in 1987, but was limited in power, suffered from selection bias and eventually showed no difference in mortality between study arms [Sandblom et al. 2004].

Two major randomized controlled trials commenced in the early 1990s: the European Randomized Study of Screening for Prostate Cancer (ERSPC) as a multicenter trial in eight European countries; and the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) in the USA. These trials recruited

very large numbers of subjects and great hopes were placed on them for resolving the controversy over PC screening (Table 2.2).

The ERSPC trial is in practice a multicenter trial consisting of several screening trials each of which has its own criteria regarding screening interval, the threshold for a positive screening result, type of recruitment and age of recruits (Table 2.3). The PLCO trial is instead a more uniform trial in 10 centers in the USA. Both trials assessed prostate cancer specific mortality as the main end-point, in addition to quality of life and cost-effectiveness assessment.

The interim mortality results from these trials were published in March 2009. The PLCO trial showed no significant difference in PC mortality (RR 1.13; 95% CI, 0.75-1.70) [Andriole et al. 2009], whereas in the ERSPC trial a relative reduction of 20% was observed (RR 0.80; 95% CI 0.65-0.98) [Schröder et al. 2009]. The ERSPC trial for the first time showed conclusive evidence that screening with PSA can reduce mortality, but 1,410 men would have to be screened and 48 PCs treated to prevent one death. In addition, in both trials screening resulted in overdiagnosis (70% in the ERSPC and 22% in the PLCO trial).

At the beginning of 2012, both trials reported results after extended follow-up. In the ERSPC trial, at 11 years of follow-up, the relative risk reduction for PC death was 21% (RR 0.79, 95% CI 0.68-0.91), and in the PLCO trial at 13 years of follow-up the RR was 1.09 (95% CI 0.87-1.36) [Schröder et al. 2012; Andriole et al. 2012].

That the PLCO trial did not show a difference in mortality has been explained by extensive contamination of the control arm, since a third of the men in it had undergone DRE and PSA testing in the past three years at the beginning of the study. Moreover, half of the control arm men underwent PSA testing during the trial. Therefore the difference in PC incidence between the trial arms was much smaller than in the ERSPC trial. Due to this major diluting effect it is likely that the PLCO trial will not yield a significant difference even with longer follow-up. The strength of the PLCO trial was, however, that the level of contamination was evaluated, whereas for the ERSPC trial there are only fragmentary estimates.

Table 2.2. Characteristics of randomized trials assessing prostate cancer mortality.

	No. of men	Age group	Screening interval (years)	Follow-up (years, median)	No. of PC deaths	RR (95% CI)
Québec trial (Labrie et al, 2004)	46,193	45-80	1	7.93	84	1.09 (0.82-1.43)
PLCO (Andriole et al, 2012)	76,693	55-74	1	11.5	174	1.13 (0.75-1.70)
ERSPC (Schröder et al, 2012)	182,160	50-74	2-7	11	761	0.79 (0.68-0.91)
Gothenburg trial (Hugosson et al, 2010)	19,904	50-64	2	14	122	0.56 (0.39-0.82)

Table 2.3. Characteristics of the screening protocols in the European Randomized Study of Screening for Prostate Cancer (ERSPC) and the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO).

	Total number of men	Interval, years	PSA cutoff, ng/ml	Age at entry (years)	Recruitment	Screening began in
ERSPC centers						
Belgium [1]	8,562	4-7	3.0	55-75	volunteer	1991
Finland [2]	80,379	4	4.0	55-67	population	1996
France [3]	84,781	2	3.0	55-69	population	2003
Italy [4]	14,517	4	4.0	55-71	population	1996
Netherlands [5]	34,833	4	3.0/4.0	55-75	volunteer	1993
Spain [3]	2,197	4	3.0	45-70	volunteer	1993
Sweden [3]	19,904	2	3.0	50-64	population	1994
Switzerland [6]	9,903	4	3.0	55-70	population	1998
Total	255,076					
PLCO (USA) [7]	76,693	1	4.0	55-74	volunteer	1993

1 Ancillary test: DRE
2 Ancillary test for PSA 3.0-3.9 ng/ml: DRE in 1996-1998; free/total PSA ratio from 1999 onwards
3 No ancillary test
4 Ancillary test for PSA 2.5-4.0 ng/ml: DRE and TRUS. Biopsy for all with PSA ≥4.0 ng/ml
5 PSA cutoff 4.0 ng/ml was lowered to 3.0 ng/ml in May 1997
6 Ancillary test: free/total PSA ratio
7 In addition to PSA, an annual DRE for the first 4 years

The Swedish component of the ERSPC trial later published the mortality results separately, and showed that in the Swedish center the mortality reduction was even greater than in the joint ERSPC analysis: 44% (RR 0.56; 95% CI 0.39–0.82; 50/10,000 vs. 90/10,000) [Hugosson et al. 2010]. Possible reasons for the larger cancer-specific mortality reduction in Sweden include that the subjects were younger than in the ERSPC in general, the screening interval was shorter (2 years), contamination was estimated to be very low at the beginning of the trial (3%), and longer follow-up. Nevertheless, no effect was observed on all-cause mortality, and overdiagnosis was again a major adverse effect (64%).

Since the publication of these new trial results two reviews and meta-analyses have been published, and the pooled data suggest that PC screening does not reduce cancer-specific or all-cause mortality, but results in worrying amounts of overdiagnosis [Djulbegovic et al. 2010; Ilic et al. 2011]. However, due to the heterogeneity of the screening trials, a meta-analytic approach is so far not necessarily powerful enough to assess the usefulness of PC screening.

Regardless of the fact that screening asymptomatic men with PSA has not yet been comprehensively evaluated, some authorities still recommend PSA testing in asymptomatic men. The European Association of Urology guidelines state that widespread PC screening is not appropriate, but opportunistic screening should be offered to well-informed men (with no statement to the age of the patient) [Heidenreich et al. 2011]. The American Urological Association guidelines do not recommend mass screening either, but suggest a baseline PSA at age 50 (or at age 40 if the patient has a family history of PC or is African American), and further testing annually or with longer intervals if the baseline PSA is very low, but not if the life expectancy is less than 10 years [American Urological Association 2009]. The Guidelines of the American Cancer Society are essentially on a par with the American Urological Association guidelines, but further recommend annual screening for men with PSA >2.5 ng/ml, biennially for men with PSA <2.5 ng/ml [American Cancer Society 2011]. The U.S. Preventive Services Task Force concludes that the current evidence is insufficient for or against PC screening, and advises against PSA-based screening in all age groups (previously only against screening in men aged over

75 years) [Moyer et al. 2012]. The Finnish Urological Association currently does not recommend systematic screening for PC, but states that PSA testing should not be withheld from symptomatic patients [Käypä hoito: Eturauhassyöpä 2007].

3. AIMS OF THE STUDY

The objective of this dissertation was to evaluate various aspects of prostate cancer screening in the Finnish component of the European Randomized Study of Screening for Prostate cancer. The specific aims were:

i) to assess the extent to which false-positive screening results occurred in repeated prostate cancer screening in the Finnish (**I**) and in the whole ERSPC (**II**) trial.

ii) to determine the possible risk for future prostate cancer, another false-positive screening result and subsequent noncompliance in screening in men with false-positive screening results (**I**, **II**).

iii) to establish how screening affects the incidence of prostate cancer – especially advanced cancer – in Finland (**III**).

iv) to determine how screening affects prostate cancer mortality in Finland, and which factors contribute most to the group of screening failures (i.e. men who die of prostate cancer in spite of screening) (**IV**).

4. MATERIALS AND METHODS

The Finnish Prostate Cancer Screening Trial is one component of ERSPC trial (European Randomized Study of Screening for Prostate Cancer), a multicenter trial. The Finnish trial was commenced in 1996, and is a population-based randomized screening trial. The data from the Finnish trial were used in three articles in this dissertation (**I, III, IV**), and the data from four other ERSPC centers (Belgium, Italy, the Netherlands and Sweden) in addition to the Finnish data were used in one article (**II**).

4.1 Study population

4.1.1 The Finnish trial

The Finnish trial included 80,144 men born 1929-1944 in the Helsinki and Tampere metropolitan areas. In the period 1996-1999, a total of 8,000 men aged 55, 59, 63 or 67 years were annually randomly assigned to the screening arm (Table 4.1), while the remainder of the birth cohort formed the control arm. This resulted in two study arms with a ratio of approximately 1.5 : 1 (control arm : screening arm). The study population was identified from the Finnish Population Register Centre. Men with a previous diagnosis of PC were excluded (these men were identified from the Finnish Cancer Registry). The men in the control arm were not contacted.

The men in the screening arm were sent a letter of invitation, a brief overview of the trial, a questionnaire about urological symptoms, previous PSA tests and family history of PC, and an informed consent form.

Table 4.1. Relation of year of birth to randomization year and screening protocol.

	Age at randomization				
Randomization year	55 years	59 years	63 years	67 years	Re-invited in
1996	1941	1937	1933	1929	2000, 2004
1997	1942	1938	1933	1930	2001, 2005
1998	1943	1939	1933	1931	2002, 2006
1999	1944	1940	1933	1932	2003, 2007

Because of logistic difficulties there were 1,671 men in the first round who were never actually invited for screening. These men were included in most analyses according to the intention-to-screen principle.

4.1.2 Other ERSPC centers

In one article (**II**) data from four other ERSPC centers were used in addition to the Finnish center (Belgium, Italy, the Netherlands and Sweden). France, Spain and Switzerland were excluded from this study because these centers had fewer than three complete rounds. In Belgium and the Netherlands men were randomized after informed consent was obtained due to legislative reasons. In Italy and Sweden men were randomized to the screening and control arms prior to informed consent and only men in the screening arm were contacted (as in Finland).

4.2 Screening protocol (intervention)

4.2.1 The Finnish trial

The men in the screening arm were invited to a local cancer society clinic for the screening test, i.e. a blood sample to determinine serum PSA concentration. Men with PSA ≥4.0 ng/ml were referred to a local urological clinic for diagnostic examinations, including DRE, TRUS and biopsy. Initially a sextant biopsy was used, but from 2002 onwards 10-12 biopsy cores were adopted. Men with PSA 3.0-3.99 ng/ml were referred to an additional test, which in 1996-1998 was DRE and F/T PSA ratio from 1999 onwards. Men with a suspicious DRE or F/T PSA ratio <16% were also referred to diagnostic examinations.

The men in the screening arm were re-invited to the second and third screening rounds four and eight years after the first screen (Table 4.1), regardless of whether they had participated previously or not. The men who had been diagnosed with PC or had emigrated from the study area were no longer re-invited. The men who were 67 years of age at the beginning of the study were not invited to the third round – at which point they would have been 75 years of age (i.e. they were screened only twice).

All the laboratory analyses were carried out at the Department of Clinical Chemistry, Helsinki University Hospital. The serum concentrations of total PSA were analyzed by both Hybritech Tandem-E (Beckman Coulter, Brea, California, USA) and Wallac Delfia (Wallac, Turku, Finland) assays. The free/total PSA ratio was determined with the Wallac ProStatus free/total PSA assay (Wallac).

The men in the control arm received no systematic interventions.

4.2.2 Other ERSPC centers

There was some variation between the centers in the screening protocol (Table 2.3), mainly in the mode of recruitment, the screening interval, PSA threshold, and the age of the men screened.

Men who chose not to participate were re-invited to the following rounds (except in the Netherlands). Due to lack of funding, the first screening interval in Belgium was extended to seven years. The Swedish center used biennial screening and thus had up to six screening rounds. In the Swedish center, the men with PSA <1.0 ng/ml in round 2 were not invited to round 3, but were subsequently invited to rounds 4-6.

4.3 Diagnostics and follow-up

PC diagnosis was based on histopathological evaluation by a pathologist, as was determination of Gleason score. TNM staging was performed by the attending clinician. Cancers with one or more of the following characteristics were regarded as advanced: T3-4, N1 or M1. Aggressive cancers were T3-4, N1, M1 or Gleason score ≥8. The original Gleason scores were used.

According to the ERSPC definition, a screen-detected PC needed to be diagnosed within 12 months from a positive screening result. An interval cancer was defined as a cancer diagnosed within the screening interval after a negative screening result. Cancers diagnosed more than 12 months but less than four years after a positive screen were classified as early recall cancers (**III**). A positive screening result that was not followed by a PC diagnosis within 12 months was labeled an FP result. Men who were not biopsied according to the protocol were not regarded as FP.

Incident PCs were identified from local or national cancer registries. The primary treatment data were obtained from the hospital records, and included radical prostatectomy, radiation therapy (external beam radiation or brachytherapy), endocrine therapy (LHRH antagonist, antiandrogen or both) or observation (either watchful waiting or active surveillance).

Information on causes of death was extracted from population registries. Cause of death committees were established in the participating centers to validate the official causes of death. The ERSPC data were maintained from a central database located in England, to which all centers uploaded their data

biannually. The Finnish database was maintained by the Finnish Cancer Registry in Helsinki.

4.4 Data analysis and statistics

In all papers, when simple proportions and risk ratios were evaluated, their 95% CIs were calculated using basic standard error formulae. A generalized linear model for binomial distribution with a logarithmic link function was used to calculate age-adjusted RRs and their CIs (**I**, **II**). Direct age-standardization for the prevalences of results was performed using the entire study population as a reference group (**II**). Spearman's correlation coefficient for proportion of PCs and FP results was calculated by summing all rounds and centers (**II**). Two-sided Pearson Chi-square test was used to compare treatment modalities between trial arms (e.g. radical prostatectomy vs. other treatment modality) (**IV**).

Cumulative incidence of PC in the screening and control arms was calculated by dividing the number of cancer cases by the number of men in each arm (**III**). Cumulative hazard of PC was estimated using the Nelson-Aalen method (**III, IV**) [Nelson 1972; Aalen 1978]. Cox regression was used to calculate incidence rate ratios (IRR) and their statistical significance (**III**).

In the fourth paper, hazard ratios (HR) were estimated for PC and all-cause mortality for the screening arm relative to the control arm using Cox proportional model (**IV**). The selection bias due to nonparticipation in the comparison between the screened men and the control arm was corrected [Cuzick et al. 1997], and the 95% confidence intervals for Cuzick-corrected RR were calculated by simulation (with 10,000 repetitions using statistical software R), under the assumption that the numbers of deaths follow a Poisson distribution (**IV**). To estimate whether or not biopsying all men with PSA ≥3.0 ng/ml could have improved our mortality results, we omitted from the SA all those screen-negative men we potentially could have prevented from dying of PC with a lower PSA threshold. Hence, we removed from the SA all those screen-

negative men who had at least once a PSA level of 3.0–3.99 ng/ml and a subsequent PC diagnosis (whether or not they died of it) (**IV**). To estimate the contribution of interval cancers to mortality, all men with interval cancers were removed from the SA, which yielded the maximum mortality reduction effect if there had been no interval cancers (**IV**).

All statistical analyses were performed using Stata 8.2 (StataCorp, College Station, Texas, USA).

4.5 Ethical considerations

The Finnish Prostate Cancer Screening Trial protocol was approved by Helsinki and Tampere University Hospital ethics committees. Permission to use cancer registry data was obtained from the Research and Development Center for Welfare and Health (STAKES, currently part of the National Institute of Health and Welfare). Under this permission, we were able to seek information on PCs from the medical records of the men in the control arm or the non-consenting men in the screening arm. The consenting men in the screening arm gave their written consent.

In other ERSPC centers (**II**), the study protocols were reviewed and approved by the respective ethics committees in each country.

5. RESULTS

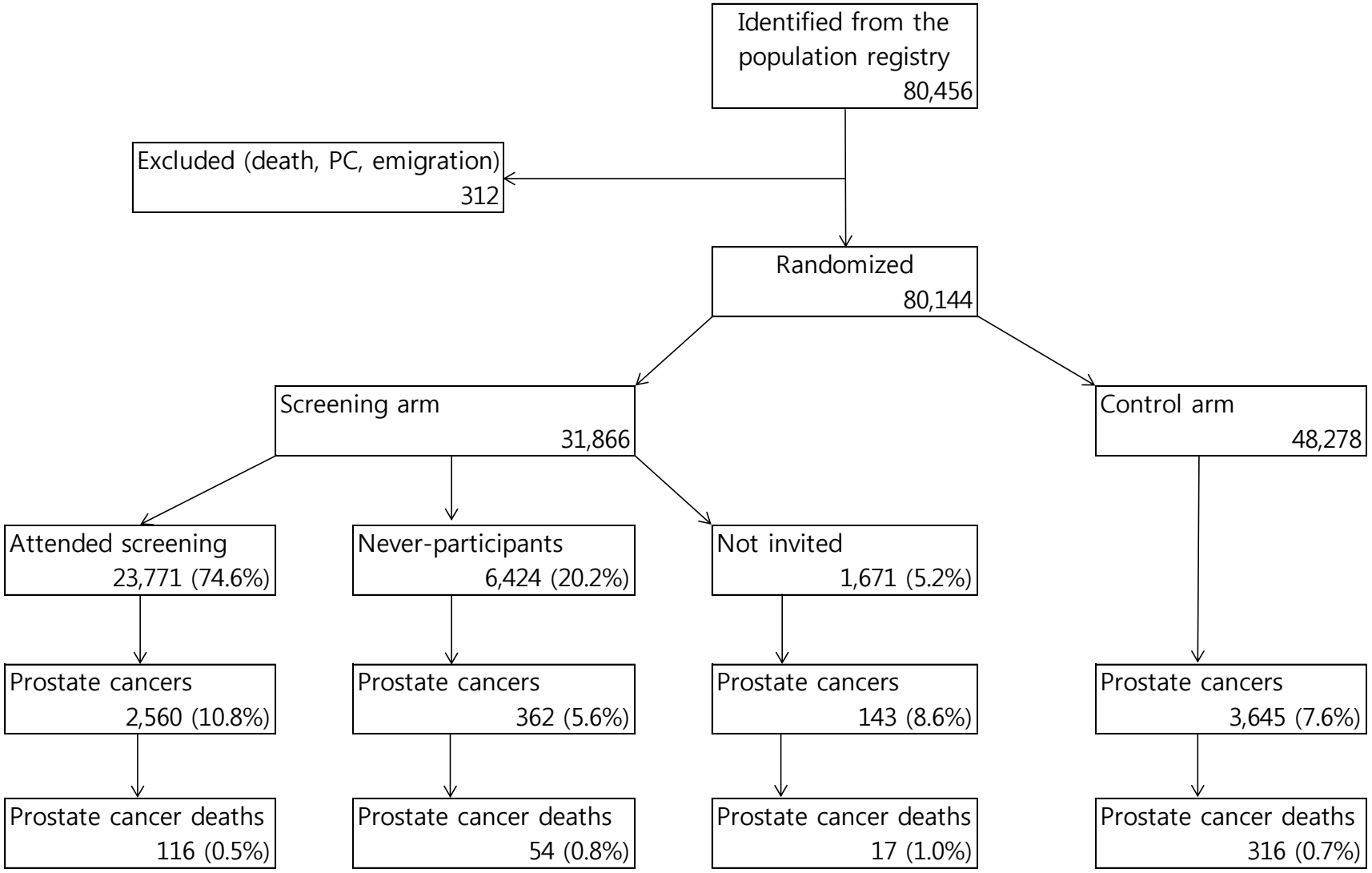
In the Finnish trial, there were 31,866 men in the screening arm and 48,278 men in the control arm (Figure 5.1). The mean follow-up time was 9.2 years (median 10.0 years) in both arms in Study **III** (common closing date December 31, 2007) and 11.9 years (median 13.0 years) in Study **IV** (common closing date December 31st, 2010). Because of the randomized design, the age distribution was similar in both arms (median 58.7 years at entry in both arms). Age proportions at entry in the SA and CA, respectively, were 55 years: 32.9% vs. 33.0%; 59 years: 26.2% vs. 26.3%; 63 years: 21.6% vs. 21.5%; 67 years: 19.2% vs. 19.2%.

Of the men in the screening arm, 74.6% participated at least once (N=23,771/31,866). In the first round, 68.8% of invited men attended screening, 70.9% in the second round and 69.5% in the third round. Altogether 10,327 men (52.1% of those invited to all rounds) attended all three screening rounds.

5.1 False-positive screening results

In all the five centers analyzed for false-positive screening results, there were altogether 61,604 men who were screened at least once. Of these, 22,068 (35.8%) men participated in all rounds. Altogether 4,733 PCs were detected by screening. This resulted in a detection probability of 3.4% (95% CI 3.2-3.5) in the first, 3.4% (3.2-3.5) in the second and 3.6% (3.4-3.8) in the third round (Table 5.1).

Figure 5.1. Flow chart of the Finnish Prostate Cancer Screening Trial.



In all centers 10.2% of men had an FP result (95% CI 10.0-10.5; specificity 89.8%) in the first round, 11.0% (10.7-11.3; specificity 89.0%) in the second round and 11.1% (10.7-11.5; specificity 88.9%) in the third round. There was substantial variation between the centers, as the proportion of FP results was lowest in Italy and highest in the Netherlands (Table 5.1). Higher age tended to increase the proportion of FP results and PCs, and there was again considerable variation (in FP results from 3.5 to 20.6%) (Table 5.2). Spearman's correlation coefficient was 0.78 ($p < 0.001$) between the proportion of prostate cancers and FP results per round.

Altogether 10,972 men (17.8% of those who were screened at least once) had at least one FP result (Table 5.3). The proportion of men with FP result(s) varied from 10.5% (Italy) to 26.1% (Netherlands). Most men (74.7%) with FP result(s) had only one such result. Of the 22,068 men who participated in all (three to six) rounds, 19.0% had at least one FP result.

An FP result increased the risk for a subsequent FP result, with approximately 50% of the men having another if they participated in the following round (Table 5.4). A negative screen resulted in an FP result in the next round in 6.2 – 7.7% of men (RR 6.5 – 8.6 by round); the RR relative to initially screen-negative men varied 2.5 – 15.0 by center. A quarter of the men with FP results chose not to participate in the subsequent screening round. This risk was 1.6-fold following the first round and 1.5-fold after the second screen compared to the screen-negative men. The RRs for nonparticipation varied by center from 0.7 (Sweden) to 2.1 (Finland).

The risk of a screen-detected PC in the next round following a previous FP result was 10%, which was roughly four-fold compared to men with a negative screen in the previous round (Table 5.4). The PPV was highest (22.0% in the first, 22.7% in the second and 25.0% third round) among those screen-positive men who had not had an FP result before. The PPV for those who had been biopsied (but not diagnosed with PC) at a previous round was 14.3% and 13.6% in the second and third rounds.

Altogether 681 cancers were screen-detected following an FP result. Of those, most (92.8%; N=632) were not aggressive, but 6.3% (N=43) were aggressive (missing information for 0.9% or six cases). Of the 1,725 PCs following a screen-

Table 5.1. Number and proportion of screening results in the five centers of ERSPC during 3-6 rounds.

	Participation proportion, N (%)	Screen-negative, N (%)	Screen-positive, N (%)		
			Not biopsied	FP	PC
All centers					
Round 1	56,064/72,210 (77.6)	47,461 (84.7)	985 (1.8)	5,722 (10.2)	1,896 (3.4)
Round 2	42,884/61,003 (70.3)	35,711 (83.3)	992 (2.3)	4,732 (11.0)	1,449 (3.4)
Round 3	27,835/42,248 (65.9)	22,929 (82.4)	819 (2.9)	3,090 (11.1)	997 (3.6)
Belgium					
Round 1	4,562/5,178 (88.1)	3,916 (88.1*)	214 (3.6*)	325 (6.1*)	107 (2.2*)
Round 2	1,987/3,430 (57.9)	1,550 (80.9*)	99 (5.8*)	237 (9.4*)	101 (3.8*)
Round 3	718/1,336 (53.7)	593 (85.0*)	49 (5.2*)	62 (8.5*)	14 (1.4*)
Finland					
Round 1	20,789/30,197 (68.8)	18,812 (90.0*)	102 (0.6*)	1,332 (7.0*)	543 (2.8*)
Round 2	18,613/26,324 (70.7)	16,309 (86.0*)	224 (1.4*)	1,467 (8.8*)	613 (3.8*)
Round 3	12,739/18,376 (69.3)	11,095 (86.3*)	198 (1.7*)	978 (8.2*)	468 (3.9*)
Italy					
Round 1	4,908/5,696 (86.2)	4,300 (88.0*)	142 (3.3*)	377 (7.1*)	89 (1.7*)
Round 2	4,499/5,607 (80.2)	3,942 (87.9*)	217 (4.7*)	267 (5.9*)	73 (1.5*)
Round 3	3,292/5,533 (59.5)	2,844 (86.1*)	269 (8.5*)	145 (4.4*)	34 (1.0*)
Netherlands					
Round 1	19,950/21,175 (94.2)	15,240 (79.2*)	470 (1.9*)	3,225 (14.6*)	1,015 (4.3*)
Round 2	12,525/16,163 (77.5)	9,259 (74.1*)	355 (2.8*)	2,360 (18.7*)	551 (4.4*)
Round 3	7,711/9,799 (78.7)	5,848 (75.8*)	217 (2.8*)	1,326 (17.2*)	320 (4.1*)
Sweden					
Round 1	5,855/9,964 (58.8)	5,193 (85.1*)	57 (1.1*)	463 (10.6*)	142 (3.3*)
Round 2	5,260/9,479 (55.5)	4,651 (87.7*)	97 (2.0*)	401 (8.0*)	111 (2.3*)
Round 3	3,375/7,204 (46.8)	2,549 (76.8*)	86 (2.4*)	579 (16.5*)	161 (4.4*)
Round 4	4,622/7,851 (58.9)	3,888 (84.1)	105 (2.3)	496 (10.7)	133 (2.9)
Round 5	4,114/6,674 (61.6)	3,499 (85.1)	69 (1.7)	435 (10.6)	111 (2.7)
Round 6	3,475/5,688 (61.1)	2,773 (79.8)	88 (2.5)	467 (13.4)	147 (4.2)

Proportions marked with an asterisk (*) are age-standardized to the mean age distribution of the round in all centers.

Table 5.2. Age-stratified aggregate results from three rounds of screening.

		Age at screen, years				
		<55 (%)	**55 - 59 (%)**	**60 - 64 (%)**	**65 - 69 (%)**	**≥70 (%)**
Round 1						
	Screen-negative	4,579 (94.9)	18,299 (90.7)	12,614 (83.9)	9,163 (77.7)	2,806 (66.4)
	Not biopsied	21 (0.4)	167 (0.8)	260 (1.7)	293 (2.5)	244 (5.8)
	False-positive	171 (3.5)	1,299 (6.4)	1,674 (11.1)	1,706 (14.5)	872 (20.6)
	Screen-detected PC	55 (1.1)	414 (2.1)	491 (3.3)	634 (5.4)	302 (7.1)
	Total	4,826	20,179	15,039	11,796	4,224
Round 2						
	Screen-negative	792 (94.9)	8,402 (91.4)	11,746 (84.9)	8,741 (79.4)	6,030 (75.3)
	Not biopsied	7 (0.8)	82 (0.9)	278 (2.0)	326 (3.0)	299 (3.7)
	False-positive	34 (4.1)	552 (6.0)	1,402 (10.1)	1,470 (13.4)	1,274 (15.9)
	Screen-detected PC	2 (0.2)	160 (1.7)	410 (3.0)	473 (4.3)	404 (5.0)
	Total	835	9,196	13,836	11,010	8,007
Round 3						
	Screen-negative	38 (84.4)	910 (82.2)	7,450 (86.1)	8,612 (81.9)	5,919 (78.7)
	Not biopsied	0 (0.0)	16 (1.4)	166 (1.9)	334 (3.2)	303 (4.0)
	False-positive	6 (13.3)	141 (12.7)	766 (8.9)	1,162 (11.1)	1,015 (13.5)
	Screen-detected PC	1 (2.2)	40 (3.6)	269 (3.1)	401 (3.8)	286 (3.8)
	Total	45	1,107	8,651	10,509	7,523

Table 5.3. Prevalence of false-positive screening results in five centers of the ERSPC trial.

	Men participating at least once, N	Men with FP(s), % (N)*	1 FP, % (N)	2 FPs, % (N)	3 FPs, % (N)	Men participating every round, N	Men with FP(s), N (%)*
All centers	61,604	17.8 (10,972)	74.7 (7,752)	20.1 (2,089)	5.2 (538)	22,068	19.0 (4,186)
Belgium	4,677	11.0 (569)	90.7 (516)	9.0 (51)	0.4 (2)	584	15.6 (105)
Finland	23,771	13.0 (2,934)	75.2 (2,207)	20.8 (611)	4.0 (116)	10,326	11.9 (1,184)
Italy	5,696	10.5 (635)	78.7 (500)	18.3 (116)	3.0 (19)	2,597	9.0 (286)
Netherlands	19,950	26.1 (5,266)	74.3 (3,912)	20.2 (1,063)	5.5 (291)	7,711	27.8 (2,228)
Sweden**	7,510	22.3 (1,568)	55.9 (876)	22.1 (347)	12.4 (195)	850	44.9 (383)
						2,112***	20.8 (442)***
			4 FPs %, (N)	**5 FPs (%)**	**6 FPs (%)**		
Sweden**			5.2 (81)	3.3 (52)	1.1 (17)		

* Age-standardized proportion

** Sweden has six screening rounds

*** Excluding the third round of the Swedish trial (the men with PSA <1.0 ng/ml in round 2 were not invited to round 3, but were subsequently invited to rounds 4-6)

Table 5.4. Risks for subsequent round PC, FP and non-participation after previous round FP result vs. negative screening result.

	Risk for non-participation after FP result, %	Risk for non-participation after negative screen, %	RR (95 % CI)	Risk for FP after FP result, %	Risk for FP after negative screen, %	RR (95 % CI)	Risk for PC after FP result, %	Risk for PC after negative screen, %	RR (95 % CI)
All centers									
R1 / R2	26.8	17.2	1.6 (1.5-1.6)	50.1	7.7	6.5 (6.2-6.8)	10.0	2.7	3.7 (3.3 - 4.2)
R2 / R3	27.8	18.6	1.5 (1.4-1.6)	53.0	6.2	8.5 (8.0-9.0)	10.0	2.6	3.9 (3.3 - 4.5)
Belgium									
R1 / R2	49.0	36.5	1.3 (1.2-1.6)	26.7	10.6	2.5 (1.8-3.6)	7.9	4.6	1.7 (0.9 - 3.4)
R2 / R3	41.3	20.6	2.0 (1.5-2.6)	39.1	4.8	8.1 (5.0-13.1)	7.8	1.3	6.0 (2.0 - 18.3)
Finland									
R1 / R2	23.6	11.6	2.0 (1.8-2.3)	53.4	5.3	10.1 (9.2-11.0)	14.7	2.5	6.0 (4.9 - 7.2)
R2 / R3	27.3	13.3	2.1 (1.8-2.3)	49.8	4.6	10.7 (9.6-12.0)	11.8	2.8	4.2 (3.3 - 5.3)
Italy									
R1 / R2	31.8	17.2	1.8 (1.6-2.2)	37.4	3.3	11.2 (8.8-14.2)	7.0	1.0	6.7 (3.9 - 11.7)
R2 / R3	39.3	32.3	1.2 (1.0-1.4)	25.3	2.6	9.6 (6.8-13.7)	1.2	0.8	1.6 (0.4 - 6.6)
Netherlands									
R1 / R2	26.2	21.0	1.2 (1.2-1.3)	51.9	13.6	3.8 (3.6-4.1)	7.2	3.9	1.8 (1.5 - 2.2)
R2 / R3	28.2	18.7	1.5 (1.4-1.6)	55.2	9.2	6.0 (5.5-6.6)	9.9	2.9	3.4 (2.8 - 4.3)
Sweden									
R1 / R2	23.8	15.7	1.5 (1.3-1.8)	49.1	4.2	11.8 (9.9-14.1)	14.7	1.0	14.4 (9.8 - 21.3)
R2 / R3	15.1	22.5	0.7 (0.5-0.9)	66.9	10.8	6.2 (5.4-7.2)	11.7	3.3	3.5 (2.4 - 5.1)
R3 / R4	15.4	17.7	0.9 (0.7-1.1)	58.5	9.4	6.2 (5.2-7.4)	8.9	3.3	2.7 (1.8 - 4.1)
R4 / R5	14.5	10.2	1.4 (1.1-1.8)	57.8	3.9	15.0 (12.3-18.2)	10.4	1.5	7.0 (4.6 - 10.6)
R5 / R6	17.3	12.0	1.4 (1.1-1.8)	63.9	7.6	8.4 (7.2-9.9)	9.9	3.1	3.2 (2.1 - 4.7)

negative result, 90.4% (N=1,560) were not aggressive and 7.8% (N=134) aggressive (1.8%, N=31 with missing information).

Had the PSA threshold been 4.0 ng/ml in all centers, the proportion of FP results would have decreased from 17.8% to 11.7% (10,972 vs. 7,182). However, fewer cancers would have been detected: 3,481 instead of 4,733 (91.5% of these missed cancers would have been non-aggressive, 6.7% aggressive, 1.8% unknown). In Belgium, the proportion of FP results would have been 6.4% (instead of 11.0%); in Finland 12.0% (13.0%); in Italy 9.3% (10.5%); in the Netherlands 12.4% (26.1%) and in Sweden 14.0% (22.3%).

In the Finnish trial, men who were FP in the first round had a fivefold risk of PC death during the follow-up of 11.9 years compared to screen-negative men: RR was 4.90 (0.3% vs. 1.3%; 95% CI for RR=2.83-8.48).

5.2 Cancer incidence in the Finnish trial

By the end of 2007, altogether 2,655 PCs had been detected in the SA (cumulative incidence 8.3%), and 2,796 cancers in the CA (5.8%) (**III**). In the following three years (until the end of 2010), an additional 410 cancers were detected in the SA (altogether N=3,065; cumulative incidence 9.6%) and 849 cancers in the CA (N=3,645; 7.6%).

The cancer detection proportion varied from the first round at 2.6% to the third round at 3.6% (Table 5.5). Per round, 88 – 95% of screen-positive men were biopsied according to the protocol. Altogether 222 interval cancers were detected, of which 15.8% were advanced at the time of diagnosis (Table 5.6). The cumulative incidence of interval cancers was 0.70%. Of 267 early recall cancers (diagnosed within 1-4 years from a positive screen), 6.4% (N=17) were advanced at the time of diagnosis. The men who did not participate in screening had altogether 567 cancers (26% advanced at the time of diagnosis). Of the screen-detected cancers (N=1,612) 8.6% were advanced at the time of diagnosis, whereas in the control arm (N=2,806) 24.4% were advanced.

During the follow-up to the end of 2010, the incidence of PC was 8.5/1,000 person-years in the screening arm and 6.3/1,000 person-years in the control

arm (incidence rate ratio IRR = 1.30; 95% CI 1.24-1.36, p <0.001). The Nelson-Aalen cumulative hazard estimates were greater for the men in the screening arm (Figure 5.2). The incidence of localized PC was 7.3/1,000 person-years in the screening arm and 4.9/1,000 person-years in the control arm (IRR = 1.48; 95% CI 1.40-1.56). The incidence of advanced PC was 1.2/1,000 person-years in the screening arm and 1.6/1,000 person-years in the control men (IRR = 0.74; 95% CI 0.66-0.83). The Nelson-Aalen cumulative hazard estimates of advanced cancer began to diverge after approximately five years of follow-up, and the difference remained at a steady level after 10 years of follow-up (Figure 5.3). The absolute effect on incidence of advanced PC can be expressed as number needed to (invite to) screening (NNS), which was 204 (95% CI 151-314), i.e. one avoided advanced PC case per 204 randomized men during a mean follow-up time of 11.9 years.

The cumulative incidence of low-grade (Gleason 2-6) cancers was 5.3/1,000 person-years in the SA and 3.2/1,000 person-years in the control arm (IRR = 1.69; 95% CI 1.58-1.80). In the SA, the cumulative incidence of Gleason 7 cancers was 2.0/1,000 person-years and in the CA 2.0/1,000 person-years (IRR = 0.98; 95% CI 0.89-1.07). There was no statistically significant difference in the incidence of Gleason 8-10 cancers, as the cumulative incidence was 1.0/1,000 person-years in the SA and 1.1/1,000 person-years in the CA (IRR 0.90; 95% CI 0.79-1.03).

Table 5.5. Results from the three screening intervals in the Finnish Prostate Cancer Screening Trial.

		Interval 1, N (%)	Interval 2, N (%)	Interval 3, N (%)	Interval 4, N (%)
Screening arm					
	PC (total)	906	1,073	689	397
	Participants				
	Screen-negative	18,812 (90.5)	16,309 (87.6)	11,096 (87.1)	-
	False-positive	1,331 (6.4)	1,488 (8.0)	1,000 (7.8)	-
	Not biopsied	103 (0.5)	202 (1.1)	189 (1.5)	-
	Screen-detected PC	543 (2.6)	614 (3.3)	455 (3.6)	-
	Total	20,789 (100.0)	18,613 (100.0)	12,740 (100.0)	-
	Early recall PC	113	123	31	-
	Interval PC	55	102	65	
	Non-participants	9,406	7,822	5,752	-
	PC	126	179	72	-
	Not invited	1,671	5,431	13,374	-
	PC	69	55	66	-
Control arm					
	PC (total)	893	1,338	575	839

Table 5.6. Distribution of prostate cancer stage in various subgroups of the trial.

	Round 1, N (%)	Round 2, N (%)	Round 3, N (%)
Screening arm			
Screen-detected PC			
Localized*	469 (86.4)	579 (94.3)	426 (93.6)
Advanced**	74 (13.6)	35 (5.7)	29 (6.4)
n/a	0 (0.0)	0 (0.0)	0 (0.0)
Total	543	614	455
Interval PC			
Localized	46 (83.6)	84 (82.4)	52 (80.0)
Advanced	9 (16.4)	14 (13.7)	12 (18.5)
n/a	0 (0.0)	4 (3.9)	1 (1.5)
Total	55	102	65
Early recall PC			
Localized	105 (92.9)	114 (92.7)	31 (100.0)
Advanced	8 (7.1)	9 (7.3)	0 (0.0)
n/a	0 (0.0)	0 (0.0)	0 (0.0)
Total	113	123	31
Non-participant PC			
Localized	120 (61.5)	183 (78.2)	112 (81.2)
Advanced	73 (37.4)	50 (21.4)	23 (16.7)
n/a	2 (1.0)	1 (0.4)	3 (2.2)
Total	195	234	138
Control arm PC			
Localized	635 (71.1)	1,015 (75.9)	460 (80.0)
Advanced	258 (28.9)	313 (23.4)	113 (19.7)
n/a	0 (0.0)	10 (0.8)	2 (0.4)
Total	893	1338	575

* Localized cancer = T1-2, N0 and M0
** Advanced cancer = T3-4, N1 or M1

Figure 5.2. Nelson-Aalen cumulative hazard estimates of overall prostate cancer risk in the screening and control arms.

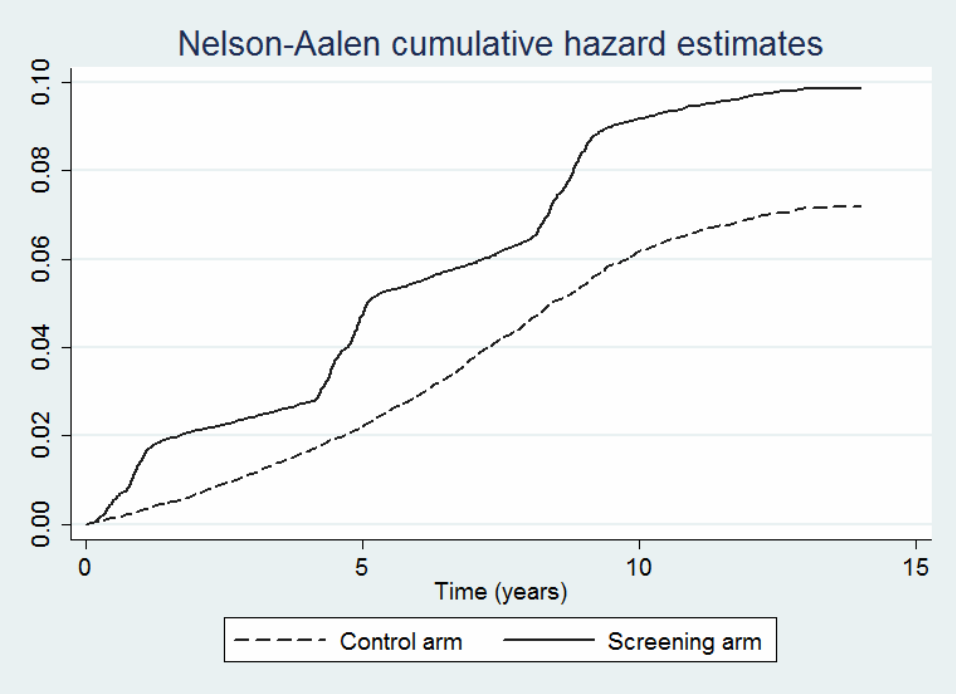
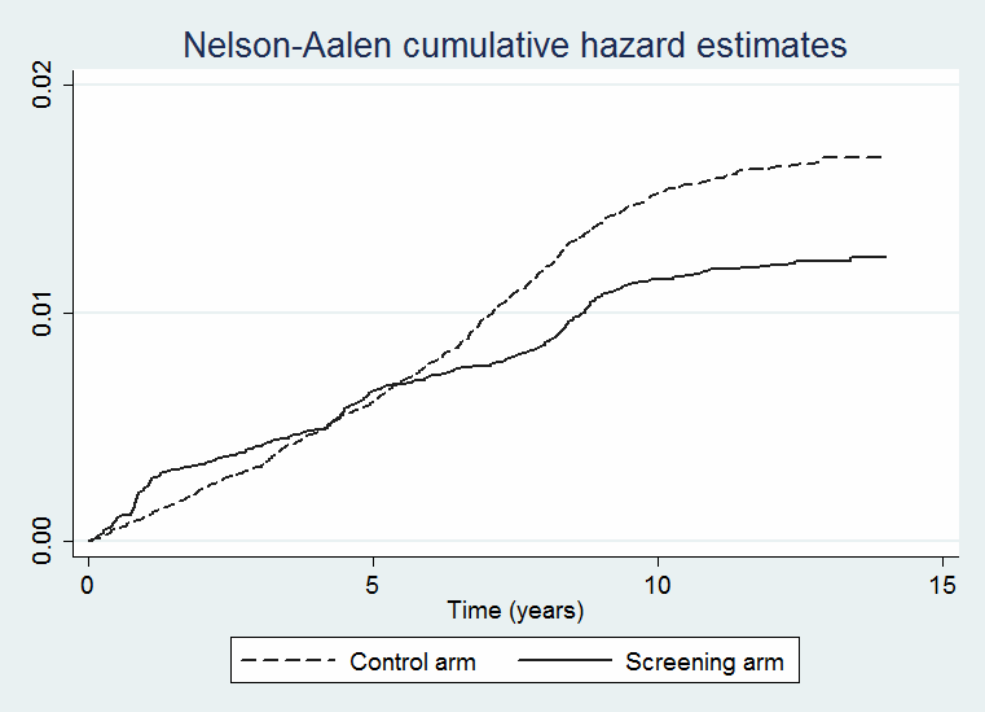


Figure 5.3. Nelson-Aalen cumulative hazard estimates of advanced prostate cancer (T3-4NxM0 or T1-4NxM1) risk in the screening and control arms.



5.3 Mortality in the Finnish trial

Altogether 187 men died of PC in the SA (cumulative mortality 59/10,000) and in the CA 316 men died (65/10,000). HR for death from PC was 0.89 (95% CI 0.75-1.07) (Figure 5.4). This corresponds to an NNS of 1,447 (95% CI 561-2,338), after a mean follow-up time of 11.9 years. The number of PCs needed to be detected to avoid a death from PC was 31 (95% CI 12-48).

The number of deaths from other causes totaled 7,731 (24.3%) in the SA and 11,697 (24.3%) in the CA. The HR for the SA relative to CA for deaths from all causes was 1.00 (95% CI 0.97-1.03). Included in the SA, there were 3,291 deaths from other causes than PC (40.7%) among the never-participants.

Of the men who died of PC in the SA, 71 (38.0%) never participated in screening (cumulative mortality 88/10,000). Altogether 47 (25.1%) men were diagnosed at their first screen, and 20 (10.7%) were due to an interval cancer. There were 8 (4.3%) men whose cancer was diagnosed 1-4 years after a test-positive but a biopsy-negative screen. As many as 23 men (12.3%) had received their cancer diagnosis after having missed the previous screening round (but had participated once earlier). Finally, 18 men (9.6%) had a screen-detected PC with either a previous screen-negative or screen-positive result.

After correcting the mortality reduction for nonparticipation by the Cuzick method, the RR was 0.84 (95% CI 0.66-1.10). To estimate the maximum effect of preventing PC death in men with PSA 3.0-3.99 ng/ml and negative screen, these men were excluded from the SA cohort. After this, an HR of 0.86 (0.71-1.03) was obtained (Table 5.7).

In a similar manner, the men with interval cancers (N = 222) were excluded from the SA to estimate how large an effect on mortality could have been achieved if all interval cancers could have been prevented. The following HR was 0.81 (95% CI 0.67-0.97).

Moderate-risk PCs were more often treated with radical prostatectomy in the SA and less often with radiation therapy compared to the CA (Table 5.8). Also, men with high-risk PC were more often treated with radical prostatectomy and less often with endocrine therapy compared to men in the CA.

Figure 5.4. Nelson-Aalen estimates of risk of dying from prostate cancer.

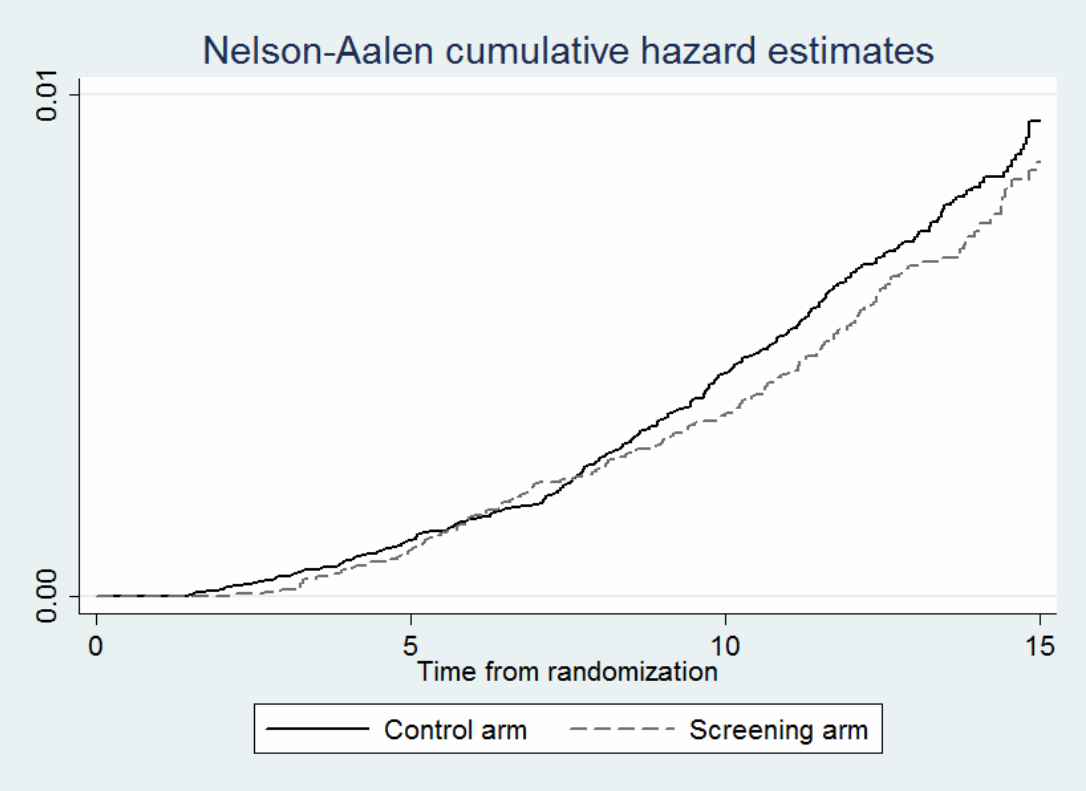


Table 5.7. Risk ratios after exclusion of specific subgroups from the screening arm.

	Screening arm			Control arm			HR (95% CI)
	No. of men	Person-years	No. of PC deaths (%)	No. of men	Person-years	No. of PC deaths (%)	
All men	31,866	380,229	187 (0.59)	48,278	575,398	316 (0.65)	0.89 (0.75-1.07)
Excluding men with PSA 3.0-3.99ng/ml	31,388	374,342	176 (0.56)	48,278	575,398	316 (0.65)	0.86 (0.71-1.03)
Excluding men with interval cancers	31,644	377,426	167 (0.53)	48,278	575,398	316 (0.65)	0.81 (0.67-0.97)

Table 5.8. Distribution of treatment modalities.

	Screening arm, N (%)	Control arm, N (%)	p for difference*
LOW RISK			
RP	312 (26.3)	242 (25.4)	p = 0.65
Radiation	320 (27.0)	278 (29.2)	p = 0.25
Endocrine	32 (2.7)	54 (5.7)	p <0.001
Expectant	498 (42.0)	354 (37.2)	p = 0.03
Total	1,162	928	
MODERATE RISK			
RP	451 (40.3)	425 (31.1)	p <0.001
Radiation	422 (37.7)	627 (45.9)	p <0.001
Endocrine	106 (9.5)	132 (9.7)	p = 0.88
Expectant	128 (11.4)	154 (11.3)	p = 0.89
Total	1,107	1,338	
HIGH RISK			
RP	100 (21.5)	95 (13.2)	p <0.001
Radiation	242 (51.9)	393 (54.8)	p = 0.33
Endocrine	100 (21.5)	189 (26.4)	p = 0.06
Expectant	13 (2.8)	17 (2.4)	p = 0.65
Total	455	694	
ADVANCED DISEASE			
RP	20 (7.1)	20 (3.4)	p = 0.02
Radiation	98 (34.6)	200 (34.2)	p = 0.91
Endocrine	154 (54.4)	337 (57.7)	p = 0.36
Expectant	3 (1.1)	7 (1.2)	p = 0.86
Total	275	564	
Low risk	=	T1 and N0/X and M0 and Gleason score ≤6 and PSA ≤10 ng/ml	
Moderate risk	=	T1-2 and N0/X and M0 and Gleason ≤7 and PSA ≤20 ng/ml	
High risk	=	T1-4 and N0/X and M0, and Gleason or PSA 20-100 ng/ml	
Advanced disease	=	T3-4 or N1 or M1 or PSA ≥100	
RP = radical prostatectomy			
*Pearson Chi-square test, two-sided.			

6. DISCUSSION

The Finnish Prostate Cancer Screening Trial and the whole ERSPC trial have both been tremendous efforts to assess the effectiveness of PC screening in reducing PC mortality. Only prospective randomized controlled trials can demonstrate this effect conclusively.

This dissertation is based on the Finnish Prostate Cancer Screening Trial and in part on data from the ERSPC as a whole. Mortality reduction, one of the main foci of this dissertation, was not observed at a statistically significant level at almost 12 years of follow-up. However, there was a clear reduction in the incidence of advanced PC in the screening arm, although at the cost of substantial overdiagnosis. In addition to the adverse effects of overdiagnosis, a large proportion of screened men were subject to a false-positive screening result.

While these observations aid in deciding whether to screen for PC or not, this dissertation did not address the cost-effectiveness or the effect on quality of life – two important factors influencing the decision on undertaking a nationwide screening program.

6.1 Assessment of the trial protocol

The trial protocol in Finland was relatively well-accepted at population level, as the participation proportion was good for a population-based trial. The proportion of screen-positive men who underwent biopsy according to the protocol (which is a prerequisite to differentiate between malignant and benign disease) was the highest among the ERSPC centers.

6.1.1 Invitation and participation

In the Finnish trial, due to logistic difficulties in arranging the screening program, there were altogether 1,671 men who were randomized to receive screening, but were never invited. Naturally, the existence of such a group of men should have been avoided. This group is included in the SA among the nonparticipants.

The participation proportion in the Finnish trial was acceptable for a population-based trial, as roughly 75% of men participated at least once. Over half of the men participated all three times, and per round the participation proportion was 70%, which compares well with the colorectal cancer screening program in Finland (participation 69 – 71% per round) [Malila et al. 2011]. In other ERSPC centers with a population-based design, the participation proportion per round varied 60 – 86% in Italy, and 56 – 62% in Sweden. Centers with a volunteer-based design naturally reached higher proportions, e.g. in the Netherlands the proportion varied from 78-94% (**II**).

In contrast to volunteer-based trials, population-based screening trials often cannot achieve complete coverage of the target population, as some people choose not to participate. The reasons for not participating may be various, such as not finding a suitable time to attend the screening test, not being fit enough to participate, or having already been tested for the target disease. An analysis from the Finnish prostate cancer screening trial revealed that the most commonly given reasons for non-participation were forgetting the invitation (51%), previous PSA testing (41%), not wanting to think of PC (39%) and regarding possible further examinations as unpleasant (28%) [Malmi et al. 2010].

When *screening efficacy* is evaluated, the people who have actually undergone screening are compared to people did not receive screening (i.e. as if the participation proportion had been 100%) [Hakama et al. 2007]. This tests whether the screening test is able to reduce morbidity or mortality, if complete attendance was achieved. In contrast, when *program efficacy* is evaluated, the nonparticipant fraction is also analyzed in the screening arm, which gives a more reliable estimate of practical efficacy (i.e. the intention-to-screen

principle). This is reflected by the fact that in the Finnish trial (**IV**), the relative mortality reduction was 11% (95% CI -7 – 25%) but after correcting the result for the bias caused by the nonparticipants the relative mortality reduction was 16% (95% CI 0.66-1.10). Although the Cuzick correction is an artificial method, it gives an estimate of the magnitude with which the nonparticipant part of the intervention group affect the observed difference in mortality.

Nonparticipant men differ from those who do participate, which is known as a "healthy screenee bias" [Zeliadt et al. 2007]. As shown by our results, both the overall mortality and prostate-cancer specific mortality were higher in the men who did not participate in screening versus the participant men or the men in the control arm. These nonparticipant men constitute a high-risk group and should always be included in analyses to reach valid conclusions when evaluating program effectiveness [Hakama et al. 2007], as we have done here.

The possible means of increasing the participation proportion if screening were to be launched nation-wide are beyond the scope of this dissertation, but may include raising awareness, ease and flexibility of participation, e.g. combining the screening test with other routine medical examinations.

6.1.2 Screening test

A choice of the screening test is always a compromise to find an optimal balance between sensitivity and specificity. PSA as a screening test is far from perfect – aggressive PCs may occur in men with very low PSA, and men with high PSA may be cancer-free. Nevertheless, it is a simple blood test, safe and inexpensive, and therefore has also several favorable characteristics for a screening test.

In the Finnish trial, a PSA threshold of 4.0 ng/ml was chosen in 1996, and an auxiliary test (DRE in 1996-1998; F/T PSA in 1999 – 2007) was offered to men with PSA 3.0 – 3.99 ng/ml. A value of 4.0 ng/ml is a relatively high cutoff point by present day standards, and many other ERSPC centers used a lower cutoff point (**II**). Had we lowered the cutoff point to 3.0 ng/ml, an unknown number of indolent PCs would have been detected. Since we did not perform an end-of-study biopsy on screen-negative men (as was done in the PCPT trial), we cannot

know exactly how many of the screen-negative men with PSA 3.0-3.99 actually harbored PC at that time. However, we can retrospectively analyze the data and observe that in men with PSA 3.0-3.99 and negative screen, altogether 11 men had PCs that resulted in death during the mean follow-up of 12 years. Biopsying all men with PSA ≥3.0 ng/ml would have meant that an extra 2,872 men would have undergone an invasive procedure. It is possible that some of the advanced cancers in men with PSA ≥3.0 ng/ml could have been detected earlier with a lower PSA threshold. Yet, this would not have made a substantial difference to mortality (**IV**), but would definitely have added adverse effects from screening to almost 3,000 men.

Furthermore, 23 advanced PCs (10% of all interval cancers) were detected outside the protocol within four years after the men had had PSA <3.0 ng/ml, meaning that lethal cancers are found even in men with low PSA levels. Similar observations have been made previously: in the PCPT trial, 2.3% of men with PSA less than 4.0 ng/ml had an aggressive PC in the end-study biopsy [Thompson et al. 2004].

The sensitivity of our screening test has been estimated previously as 0.89, which is comparable to both the Swedish (0.90) and the Dutch (0.93) ERSPC centers [Auvinen et al. 2009]. The Swedish center used biennial screening and the Dutch screened with a four-year interval; both used a PSA threshold of 3.0 ng/ml. It is likely that sensitivity could have been improved by lowering the PSA threshold in Finland too, but it is debatable whether the improved sensitivity would have been worth the loss in specificity, i.e. the increased number of FP screening results.

The auxiliary test was changed after the first three years of the trial from DRE to F/T PSA with a cutoff of 16%. While DRE still has its place in routine urological examination, it is rather cumbersome as a screening test, as it needs to be performed by a urologist, and requires another visit to the screening clinic (if DRE is performed as a supplementary test). Furthermore, DRE is dependent on the skill of the physician, and may miss potentially lethal impalpable cancers [Schröder et al. 1998]. F/T PSA, on the other hand, is a simple auxiliary laboratory test and therefore more suitable for screening purposes, especially if F/T PSA is conditionally determined only if the total PSA is e.g. 3 – 10 ng/ml. In

this PSA range, F/T PSA is useful in differentiating between malignant and benign disease [Roddam et al. 2005; Finne et al. 2008]. As published previously from the first round of the Finnish trial, F/T PSA provided slightly better specificity (93.3% vs. 91.7%) and a higher detection rate (4.8% vs. 2.9%) compared to DRE [Mäkinen et al. 2001].

6.1.3 Screening interval

Defining the screening interval is an important decision in a screening program. The optimal screening interval is influenced by both the lead time of the particular cancer and the screening test selected. If the lead time is long, a less frequent screening interval is appropriate, whereas for cancers with a short lead time (i.e. aggressive cancers) a more frequent interval is needed to detect them before they progress to an incurable stage. Clinically emerging interval cancers (not detected by the screening protocol) provide information for estimating sensitivity and optimal screening interval [Auvinen et al. 2009].

Estimates for the lead time in PC have varied between 3 and 12 years, but an analysis utilizing three models estimated lead time to be 5.4 – 6.9 years [Draisma et al. 2009], and another study from the ERSPC trial with 4.0 ng/ml cutoff determined lead time to be 6.8 years [Finne et al. 2010].

A screening interval of four years was used in most ERSPC centers; only Sweden used biennial screening and in Belgium the interval varied from two to seven years. The PLCO trial in the USA, by contrast, screened annually. It has been observed that at 10 years of follow-up there is no major difference in the cumulative incidence of interval cancers between the Swedish (0.74%) and Dutch (0.43%) centers [Roobol et al. 2007]. Similarly, our trial reported a cumulative incidence of interval cancers of 0.53% at 9.2 years of follow-up (**III**) and 0.70% at 11.9 years of follow-up (**IV**). In the Belgian center, however, the cumulative incidence of interval cancers was 3.0% at 10 years of follow-up due to the prolonged first screening interval, up to seven years [Nelen et al. 2010]. In Belgium, the incidence rate of (aggressive) interval cancers was higher after four years from the screen suggesting that the screening interval should not

exceed four years. It is possible that more frequent screening could be more effective in reducing the incidence of advanced PC and thus mortality, as a recent report comparing Dutch and Swedish trials suggests [van Leeuwen et al. 2012]. This is further supported by a simulation analysis from the Finnish trial indicating that a shorter screening interval is a more important factor affecting the relative mortality reduction than the age of onset of screening [Wu et al. 2012]. So far no analyses have been published on sensitivity or interval cancer incidence from the PLCO trial. It is currently not known which interval is best to achieve the optimal sensitivity and specificity, although the aforementioned observations suggest it could be somewhere between two and four years, but not exceeding four years.

6.1.4 Diagnostic examinations

The screen-positive men were referred to diagnostic work-up, i.e. DRE, TRUS and biopsy. However, not all screen-positive men were actually biopsied according to the protocol. In Finland, the proportion of the screen-positive men who did not undergo biopsy varied from 0.5 to 1.6% of all screened men per round, whereas in other centers the proportion of these men varied from 1.0 to 8.2% per round (**II**).

Some of these men may have contacted their regular physicians or urologists (e.g. in private clinics) for diagnostic examinations after having learned they had tested screen-positive. In the Finnish trial, we have no direct knowledge of private sector examinations or biopsies. We can, however, indirectly observe from the Finnish Cancer Registry how many of these were subsequently diagnosed with PC. During three rounds, of the 496 men who were screen-positive but did not undergo biopsy according to the protocol, only 5 (1.0%) were later diagnosed with PC (within four years of the screen). This suggests that most of these men did not undergo biopsy anywhere, because usually 25 – 30% of screen-positive men have cancer (**I**). Several reasons could lie behind refusing biopsy. For example, these men may have been distressed by the prospect of a biopsy procedure or possible cancer diagnosis. They may also

have been tested previously with PSA and possibly biopsied, which could explain why they did not want to undergo another biopsy.

In some centers (Belgium, Italy) the proportion of non-biopsied screen-positive men of all screen-positive men was as high as 30 – 60% per round (in Finland, the Netherlands and Sweden 5 – 15% per round) (**II**). Such high figures suggest fundamental problems in the execution of the screening protocol in the Italian and Belgian centers. Still, compared to the PLCO trial, which biopsied only 30-40% of screen-positive men [Grubb et al. 2008], the overall biopsy rate was much higher in these five centers of ERSPC (85-90%).

In the Finnish trial, a sextant biopsy was used until 2002, when 10-12 cores were adopted. In other ERSPC centers, a sextant biopsy was recommended and used from 1996 onwards. The change in the Finnish biopsy protocol was made because in routine clinical practice 10-12 cores had become common, and the trial wanted to maintain comparability between the study arms. The effect of increasing biopsy cores is difficult to estimate due to the aging cohort and previous screening, but comparing screening results from 2001 and 2003 showed essentially no difference in the proportion of screen-detected cancers or FP results. In theory, using more cores should increase the cancer detection rate, as sextant biopsy may miss up to 30% of PCs [Norberg et al. 1997]. Increasing cores should in turn decrease the proportion of false-positive screening results.

All PC diagnoses were based on routine assessment by pathologists both in the screening and control arms. The interobserver variability was reasonably low within the ERSPC: the concordance between Swedish and Dutch pathologists was estimated at 87% [van der Kwast, 2006]. As the ERSPC trial has been ongoing for several years, the criteria for Gleason grading have changed (see section 2.3.3.3), which poses a problem in comparing older cancer diagnoses to present-day diagnoses. Fortunately, this change is similar in both study arms and therefore is nondifferential.

6.1.5 Treatment

The PCs were treated in the same local hospitals irrespective of trial arm. This should have resulted in equal distribution of treatment modalities between trial arms, but statistically significant differences were observed, as men in the SA were more likely to receive radical prostatectomy for moderate-risk PC and less likely to be treated with radiation therapy. Also, men with high-risk PC were more often treated with radical prostatectomy and less often with endocrine therapy compared to men in the CA.

While actual treatment analyses are beyond the scope of this dissertation, certain observations may be pointed out. The difference in treatment modalities may contribute to the mortality effect if men in the intervention receive more intensive and curatively-aimed therapy. This kind of bias would magnify the mortality effect in favor of the intervention arm.

A similar dissimilarity in treatment modalities was observed previously in the eight centers of the ERSPC trial: radical prostatectomy and active surveillance were more often chosen in the SA and endocrine therapy in the CA [Wolters et al. 2010]. As discussed by Wolters and colleagues, this may reflect a screening effect: the inherent differences between the clinically emerged PC and screen-detected PC rather than actual bias in the treatment between trial arms.

6.1.6 Follow-up

In Finland information on PC diagnoses was extracted from the database of the Finnish Cancer Registry, to which medical institutions are obliged to report new cancer diagnoses. Of cancers in Finland, 99% are reported to the Finnish Cancer Registry [Teppo et al. 1994]. In addition, both the screening and control arm cancer diagnoses were confirmed from the patient records in the local hospitals. In other ERSPC centers population-based nationwide cancer registries (Sweden) or regional cancer registries (Belgium, Italy, the Netherlands) were utilized for extracting the data.

Valid information on cancers diagnosed in both study arms is vital to reach the correct conclusions on the screening effect in a trial. In a population-based screening trial, it is often easier to obtain information on the cases in the intervention arm than in the control arm (e.g. previous PSA testing, quality-of-life questionnaires). Underreporting in the control arm is likely to magnify the difference in cancer incidence between study arms, since screening usually leads to overdiagnosis. However, if screening reduces mortality, the difference in mortality is likely to be diluted.

Mortality information was obtained in Finland from the Official Statistics of Finland, and excellent agreement (κ=0.95) was observed in PC deaths between the cause of death committee of our trial and official death certificates [Mäkinen et al. 2008]. A similar analysis from the Dutch center also provided relatively good concordance (κ=0.76) with their cause of death committee and official Dutch statistics [Otto et al. 2010]. Other centers have not published results from the cause of death committees. It may be seen as a limitation that the cause of death committees were not independent of the trial.

Cause of death may be either a clinical conclusion by the clinician or a conclusion based on autopsy. Determining a cause of death, especially in elderly patients with extensive co-morbidity, is not always simple. In Finland it has been estimated that approximately 30% of the deceased undergo autopsy [Lahti and Penttilä 2001].

In a cancer screening study correct classification for the cause of death is crucial to avoid misattribution. Therefore cause of death committees should be established in any study which uses mortality as an end-point to review whether the causes of death are indeed correct. In both ERSPC and PLCO trials such committees were used [Miller et al. 2000; de Koning et al. 2003]. As discussed in Section 2.5, cancer-specific mortality is always subject to misattribution bias, and thus overall mortality was also reported in our trial.

6.2 False-positive screening results

The results of this dissertation showed that false-positive screening results are common in PC screening as they affect 11.1 – 26.4% of the men attending repeated screening, depending on the center (**II**). The men with FP results were at higher risk for future diagnosis of PC compared to men with a negative screen (**I**, **II**), and also at higher risk for dying of PC. In addition, men with FP results often have persistently high PSA levels, which often results in subsequent FP results (**I**, **II**). Furthermore, they tend to be more likely not to participate in the next screening round than men with negative screening results (**I**, **II**).

The FP results are not surprising when discussing screening for PC with PSA. As discussed, the specificity of PSA is not very high at levels 3 – 10 ng/ml, as PSA may be elevated due to a number of common benign conditions. We followed an ERSPC definition where a screen-positive cancer should be diagnosed within one year, and if no such diagnosis is made, the result is false-positive. However, this definition is not entirely straightforward, as screening tests are not 100% sensitive and specific. If the time limit is too short, (missed) cancers will be diagnosed at repeat biopsy (i.e. an FP result becomes a true positive); and if the time limit is too far away, *de novo* cancers will arise in FP men.

Although a prostatic biopsy as a procedure is well-tolerated and rarely results in complications, it is unpleasant and often painful even with the use of a local anesthetic [Mäkinen et al. 2002; Carlsson et al. 2011; Rosario et al. 2012]. Also, waiting for the results is psychologically stressful and may cause anxiety for up to a year, even after a negative biopsy result [Fowler et al. 2006].

The proportion of FP results has been estimated at 7-8% per screening round [Lafata et al. 2004; Määttänen et al. 2007]. In repeated screening, an analysis from the PLCO trial determined that after annual screening altogether four times, 10.4% of men had at least one FP screening result [Croswell et al. 2009].

We estimated a slightly higher proportion of ever-false-positive men in the Finnish trial during three screening rounds (12.5%), but subsequently large differences were observed among ERSPC centers (**II**). Moderate proportions were observed in Belgium, Italy and Finland, whereas in the Netherlands and

Sweden more than 20% of men had at least one FP result. Unfortunately, the data from Belgium and Italy were not as reliable as in the other three centers, as 30 – 60% of men with elevated PSA were not biopsied (i.e. we cannot know whether they would have had a malignant or benign biopsy finding). Screening interval and age did not appear to be associated with the prevalence of FP results. The reason for the higher prevalence in the Swedish and Dutch centers appeared to be the lower PSA cutoff point, which is a logical determinant for lower specificity. The prevalence of FP results would have been decreased from 18% to 12% (reducing a total of 3,800 FP results, 35% of all FP results) if all centers had used a PSA threshold of 4.0 ng/ml, but 1,250 cancers (26% of all cancers) would have been missed during 3-6 screening rounds(**II**).

FP men were more likely than screen-negative men to not participate in subsequent screening – an effect observed in both the PLCO trial [Ford et al. 2005] and now in our studies (**I**, **II**). This may be explained by a (potentially false) reassurance of being free from disease, or by anxiety after an unpleasant experience. Also, receiving a positive screening result without being diagnosed with cancer may erode a man's perception of the effectiveness of screening.

Men with FP results were also determined to be at higher subsequent risk for PC in the future, but this observation must be viewed with caution. It remains unknown whether this increased risk is due to a real biological susceptibility for developing de novo PC, or simply because these men were already carriers of a minimal carcinoma lesion that was missed in the original (sextant) biopsy and detected by a rebiopsy. Men with FP results receive more follow-up biopsies, as reported previously [Fowler et al. 2006] and confirmed by the Finnish trial (**I**). Due to this more intensive follow-up, the FP results add to the costs of screening [Lafata et al. 2004].

One could argue that cancers detected after an FP screening result were mostly (>90%) localized and low-grade and thus clinically insignificant. Nevertheless, 6.3% of cancers diagnosed after an FP screening result were aggressive in nature (**II**); and even after one FP screen, there was a fivefold risk of dying of PC compared to those testing screen-negative.

Men with FP results were likely to have persistently high PSA – in fact the probability of another FP screen was more than 50% if these men participated

in the next round. This is no surprise, as the most likely explanation for an elevated PSA in these men is BPH. However, considering that FP men are likely not to participate in screening although harboring cancers (some of which are lethal) the existence of this relatively large group of FP men is no trivial matter in PC screening. Unfortunately, this dissertation does not provide clear-cut answers as to how we could reduce the number of these findings.

6.3 Impact of screening on incidence and mortality

The rationale for cancer screening is to decrease cancer-specific mortality and thus, all-cause mortality. The only way to achieve this is by detecting the cancer lesions at an early stage when they are still curable, but early detection does not automatically result in lower mortality due to lead time bias (see Section 2.5). The incidence of advanced cancers was 25% lower in the SA compared to the CA, but the overall incidence of PC was higher in the screening arm due to both overdiagnosis of low-grade localized cancers and the aforementioned lead time bias (i.e. the PCs were detected earlier than in a clinical setting) (**III**). However, in the Finnish trial at 11.9 years of follow-up only a statistically non-significant 11% relative mortality reduction was observed between the two study arms (**IV**).

When a screening program is implemented, there is an incidence peak due to diagnosing prevalent cancers in the population. If there is no overdiagnosis present, the incidence should be lower just after the peak than before screening, and gradually return to the same level as before screening. Ideally, after the implementation of screening the incidence of advanced stage cancers should remain lower than without screening, which is a prerequisite for lowered mortality.

With PC, a clear peak in the incidence can be observed after each screening round in the Finnish trial (**III**). However, due to overdiagnosis, the incidence remained high throughout the trial, and the cumulative incidence was 30% higher in the screening arm than in the control arm.

PC screening results in an excess of cancer diagnoses, which has been observed in many studies. In the joint ERSPC publication, the incidence was 70% higher in the screening arm than in the control arm [Schröder et al. 2009], and even in the PLCO trial the incidence was 22% higher in the SA despite contamination in the SA and the fact that one third of the men had been tested with PSA before the trial started [Andriole et al. 2009]. In the Swedish center the screened men had a 64% higher risk for PC diagnosis compared to the nonscreened men [Hugosson et al. 2010]. While it is logical that screening results in a plethora of cancer diagnoses, these figures suggest major overdiagnosis, i.e. detection of cancers that would not have surfaced clinically. The amount of overdiagnosis cannot be directly observed from the aforementioned figures, but studies using stochastic natural disease history models have estimated that in screening, 23-42% of screen-detected cancers are overdiagnosed [Draisma et al. 2009].

Overdiagnosis and subsequent overtreatment are probably the most important adverse effects of PC screening. Overdiagnosis may result when the diagnosed cancer progresses so slowly that the patient dies from other causes before the cancer becomes symptomatic, either due to the indolent nature of the cancer or the short life expectancy of the patient [Welch and Black 2010]. It is impossible to definitely know which PCs progress and which do not, although it is becoming increasingly clear that the risk of PC-specific death is very low in the low-risk PC group [Bul et al. 2012]. Understandably, for the clinician and especially for the patient, it may be difficult to not opt for aggressive (radical) treatment of the cancer to avoid the possibility of disease progression (even if this possibility was very low). As prospective evidence accumulates on active surveillance schemes this psychological burden of expectant management is hopefully alleviated.

An important step in quantifying the adverse and beneficial quality of life effects was taken recently when Heijnsdijk and colleagues reported of a simulation model which evaluates net gain or loss of quality of life resulting from PSA screening [Heijnsdijk et al. 2012]. This theoretical model showed that adverse effects of overdiagnosis and overtreatment were the heaviest

contributors to lost quality-adjusted life years. It remains to be seen whether or not this model can aid in clinical decision-making.

Ideally, PC screening prevents deaths from PC. When the men who die of PC despite being randomized to receive screening are analyzed, two prominent groups emerge not only in the Finnish trial but also in the Swedish [Bergdahl et al. 2009] and Dutch trials [Zhu et al. 2011]. The first group is men who do not participate in screening. As discussed earlier, nonparticipants are at a higher risk for PC and dying of PC than other men in the screening arm or men in the control arm. Another large group of men who die of PC are those who are diagnosed at the first time they attend screening. These men cannot be regarded as failures in the screening protocol *an sich*, as their existence is logical when screening is commenced. If the screening is begun at an earlier age (as in the Swedish trial), the number of these men will be smaller. Yet even in the Swedish trial the men who were diagnosed at their first screen contributed a third to the number of men who died of PC despite screening at 13 years of follow-up. Roughly 10% (N=20) of the men who died of PC in the Finnish trial despite being randomized to receive screening were due to an interval cancer. Of these men, 17 had had a PSA level <3.0 ng/ml at a previous screen. It remains unknown whether these men could have been diagnosed sooner if the screening interval had been shorter, but on average these men were diagnosed with cancer 2.7 years after the PSA test, suggesting that for some of these men diagnosis could have been made if screened biennially. Whether or not this would have prevented PC deaths remains open to conjecture.

As discussed earlier, a prerequisite for mortality reduction is earlier detection of cancers by screening, i.e. a stage shift must be observed. There is evidence of such stage shift in the Finnish trial (**III**) and in other ERSPC centers [Hugosson et al. 2004; Postma et al. 2007]. In the Swedish center, the incidence of advanced cancer was shown to be almost 50% lower (0.24 vs. 0.47%, NNS = 435) in the screening arm than in the control arm [Aus et al. 2007], and in the Finnish trial, the reduction was 30% (1.0 vs. 1.4%, NNS = 270) (**III**). Yet we observed only a small non-significant mortality reduction (11%) in the Finnish trial (mortality

among screened men 59/10,000 vs. 65/10,000 in the control men). In the pooled ERSPC analysis a 20% relative reduction was seen overall (29/10,000 vs. 36/10,000) [Schröder et al. 2009], and an even larger (44%) relative effect in the Swedish center with longer follow-up (50/10,000 vs. 90/10,000) [Hugosson et al. 2010]. However, no single center has sufficient statistical power to conclusively analyze effect on mortality [de Koning et al. 2002], which is why the original ERSPC mortality publication in 2009 did not show mortality results separately by center [Schröder et al. 2009].

Nevertheless, it is curious that the relatively large reduction in the incidence of advanced cancer did not produce a more profound mortality reduction effect (11%) in the Finnish trial (**IV**). The most obvious explanation for this could be contamination in the control arm, which has not yet been thoroughly estimated in the Finnish trial. Contamination (i.e. unorganized or wild PSA testing) in the control arm would dilute any differences observed between the arms, as was the case in the PLCO trial (half of the men in the control arm received PSA testing by the fifth year of screening and a third of the men had already been tested with PSA and DRE before the study) [Andriole et al. 2009]. Conversely, the Swedish center has estimated that only 3% of men in their control arm had been tested with PSA at randomization [Hugosson et al. 2010].

A questionnaire for physicians was implemented in the Finnish trial showing that 18% of respondents admitted having systematically screened asymptomatic men with PSA in 1999 and 9% in 2007 in addition to those 70% who used PSA screening occasionally [Pogodin-Hannolainen et al. 2011]. This indicates that contamination is likely to have a diluting effect on mortality difference between the trial arms in the Finnish trial despite the fact that such a questionnaire study with a less than 50% response rate is likely to overrepresent PSA use due to selection bias.

Lack of a reliable estimate of contamination in the control arm is a major weakness in two articles of this thesis (**III**, **IV**) and in the Finnish screening trial as a whole.

The differences in the treatment modalities may have affected mortality difference in the trial arms, but these differences would be more likely to

magnify than dilute mortality difference, as the men in the SA were more likely to receive curative-aimed treatment.

Another reason for the lower mortality reduction effect could be simply the length of the follow-up period. When the trial has serious flaws (such as major contamination in the PLCO trial), the difference between the trial arms is unlikely to emerge with further follow-up. Our study (**IV**) reported mortality results at a median of 13 years of follow-up, although the Swedish analysis had only a slightly higher median of 14 years. In Sweden, the differences between the arms began to be clearer after 11 years [Hugosson et al. 2010]. It is possible that the observed reduction in advanced PCs may still turn into a significant mortality reduction with additional follow-up time in the Finnish trial.

It has been hypothesized that PSA screening may be effective in reducing mortality in younger men, as the Swedish center had younger men as participants than did other ERSPC centers. This hypothesis could be further fortified by a post hoc subgroup analysis from the PLCO trial, in which PC mortality was observed to be reduced by screening when only men with no significant comorbidities were analyzed [Crawford et al. 2011]. The validity of this *post hoc* analysis was subsequently questioned, as the criteria for comorbidities were rather loose [Bach and Vickers 2011] and the updated PLCO publication did not corroborate these results [Andriole et al. 2012]. There is no substantial evidence so far that screening younger men is a major issue in achieving effective reduction of PC mortality. While it is true that curative treatment in younger men with high-risk localized PC produces more life years gained than does the treatment of such a cancer in elderly men, it is also undeniable that younger men would have to live longer with adverse effects in case of overtreatment. Hence analyses regarding quality of life are warranted in the future to evaluate the usefulness of PC screening, especially before recommending screening of younger men (<55 years).

7. CONCLUSIONS AND FUTURE PROSPECTS

This dissertation was intended to evaluate possible disadvantages and benefits of prostate cancer screening. The only significant benefit from screening was a decreased incidence of advanced cancer, which was not reflected in an equally large reduced mortality impact during the 12 years of follow-up but may do so when the follow-up time is extended. However, several downsides were observed for the screened men: abundance of FP screening results, and also a rough estimate of overdiagnosis and overtreatment of PCs. It is likely that with a lower PSA threshold and more frequent screening a more substantial mortality effect could have been observed, but this beneficial effect should be weighed against the adverse effects of intensive screening.

PC screening is a delicate subject. Attitudes toward screening may be very different in a man who has been diagnosed early with an aggressive cancer and treated curatively compared to a man who has undergone radical treatment for an indolent cancer and who has to suffer from the adverse effects of treatment for years. Epidemiologists need to try to distance themselves from the individual patient and see the big picture. Public health choices are difficult as resources are limited and need to be utilized to gain maximum health benefit. This dissertation can only provide limited information to aid the decision-making process, as quality of life and cost-effectiveness were not evaluated.

The results from this dissertation do not support generalized population-based PC screening. It is possible that with longer follow-up the mortality reduction will improve, but even then the cost-effectiveness and quality of life factors need to be weighed against the observed benefits. Nationwide PC screening could be justified, if (in addition to general screening criteria) the

following criteria were met: 1) the screening test and further diagnostic procedures could not only differentiate better between benign and malignant disease but also between clinically insignificant and significant disease 2) active surveillance methods for overdiagnosed low-risk cancers would be effective, and more widespread and also acceptable among patients (i.e. radical treatment could be avoided thus alleviating the treatment burden on low-risk disease) 3) it would be possible to identify high-risk subgroups in the population to which to target the screening measures. Focusing future research on how to meet these criteria is needed to improve PC screening.

KIITOKSET/ACKNOWLEDGEMENTS

En elättele illuusioita siitä, että tämä väitöskirjatyö jotenkin ilmentäisi omaa erinomaisuuttani tieteen parissa. Olen toki nähnyt paljon vaivaa kirjan valmistumisen eteen, mutta kymmenet ihmiset seulontaprojektissa ovat tehneet lähes parinkymmenen vuoden aikana valtavan määrän työtä protokollan suunnittelussa, itse seulonnan järjestämisessä, aineiston ylläpitämisessä ja talouspuolesta huolehtimisessa. PSA-seulonta alkoi Suomessa, kun olin vasta ala-asteella!

Hyvä väitöskirjatyön ohjaaja on painonsa arvoinen kultaa nuorelle tutkijalle. Minua on kohdannut verraton onni saadessani tehdä väitöskirjani kahden timantinkovan ohjaajan kanssa.

Teuvo Tammela on ollut näinä vuosina innostava ja rohkaiseva ohjaaja. Hän on – jostakin käsittämättömästä varannosta – löytänyt aikaa ja energiaa välittää aina muutaman kannustavan sanan vaikka huonostikin menneen esitelmän jälkeen. Teuvon käytännönläheinen suhtautumistapa ongelmiin on saanut monet asiat järjestymään sujuvasti, ja hänen kertomuksensa kirurginuraltaan ovat olleet arvokkaita nuorelle parturismiehelle.

Anssi Auvinen on lääketieteellisen tutkimustyön helmi. Hän on vaativa ja tarkka, mutta silti aina ystävällinen, aina innostava ja energinen. Anssi korostaa tieteen etiikkaa, ja onnistuu välittämään tutkijoilleen, kuinka tärkeää on pyrkiä mieluummin tieteelliseen laatuun kuin määrään. En voisi kuvitella parempaa väitöskirjaohjaajaa.

Liisa Määttänen on ansainnut kiitokseni moneen kertaan. Olen lähettänyt lukemattoman määrän sähköposteja, aineistokyselyjä ja selvennyspyyntöjä, ja kiireistään huolimatta Liisa on auttanut minua selvittelemään massiivista aineistoa. Aivan erityisesti haluan kiittää sveitsiläisten omenoiden lähettämisestä.

Matti Hakama ja Nea Malila ovat molemmat auttaneet minua ymmärtämään tieteellisen työn tärkeintä osa-aluetta: ajattelemista. Ilta-aikaankin he jaksoivat vääntää ilmeisen loppumattomasta rautalangasta, miksi a voi olla b mutta b voi olla olematta a.

Mirja Ruutu, Paula Kujala ja Ulf-Håkan Stenman ovat osaltaan tietysti olleet vuosien ajan koko seulontaprojektin keskeisiä moottoreita, mutta myös auttaneet minua henkilökohtaisesti matkan varrella käsikirjoitusten ja analyysien kanssa.

Jani Raitanen on ystävällisesti auttanut useampaan otteeseen tilastoanalyysiongelmissa, mistä suuri kiitos.

Kiitän myös Pekka Karhusta tieteellisen tutkimuksen perusasioiden opettamisesta ensimmäisten opiskeluvuosieni aikana.

I owe an enormous thank you to all the people with the ERSPC who have worked hard for years for the project. I want to thank especially my co-authors Monique Roobol, Jonas Hugosson, Sue Moss, Stefano Ciatto, Vera Nelen, and the man behind it all, Fritz Schröder.

I am grateful to Gunnar Steineck from Karolinska Institutet who kindly reviewed this dissertation.

Kiitokseni on ansainnut myös Antti Rannikko väitöskirjaani paneutumisesta ja esitarkastamisesta.

I thank Virginia Mattila for the rapid and professional linguistic revision of the dissertation.

Kiitän sisaruksiani ja äitiäni Teijaa koulutusta arvostavan ilmapiirin luomisesta alkukodissani.

Suurimman kiitokseni olen velkaa ihastuttavalle (engl. *enchanting*) vaimolleni Suville, jota ilman olisin tuskin koskaan aloittanut väitöskirjaprojektia. Hän onnistui tartuttamaan minuunkin kiinnostuksen syöpäepidemiologiaan, ja suositteli minulle loistavaa ohjaajaa. Suvi on monella tapaa poikkeuksellinen ihminen, ja kannusti minua eteenpäin tutkimustyössäni toisinaan kehuen ja toisinaan ystävällismielisesti sättien. Viimeisimmän vuoden aikana myös tyttäremme Kaisla on tuonut mitä suurinta onnea elämäämme.

* * * * *

Tätä väitöskirjaa tuki taloudellisesti Suomen Kulttuurirahasto (Aino Valvaalan rahasto, Pirkanmaa), Suomen Syöpäjärjestöt, Suomen Lääketieteen säätiö ja Lääkäriseura Duodecim, Maud Kuistilan säätiö, Suomen Urologiyhdistys ja Tampereen yliopisto.

REFERENCES

Aalen OO. Non parametric inference for a family of counting processes. Ann Stat. 1978;6:701-26.

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

Albertsen PC. When is active surveillance the appropriate treatment for prostate cancer? Acta Oncol. 2011;50 Suppl 1:120-6.

Alcaraz A, Hammerer P, Tubaro A, Schröder FH, Castro R. Is there evidence of a relationship between benign prostatic hyperplasia and prostate cancer? Findings of a literature review. Eur Urol. 2009;55:864-73.

American Cancer Society. www.cancer.org/Cancer/ProstateCancer/MoreInformation /ProstateCancerEarlyDetection/prostate-cancer-early-detection-pdf. (Accessed October 7th, 2011.)

American Urological Association. Prostate-Specific Antigen Best Practice Statement, 2009. www.auanet.org. (Accessed October 7th, 2011.)

Andrén O, Fall K, Franzén L, Andersson SO, Johansson JE, Rubin MA. How well does the Gleason score predict prostate cancer death? A 20-year followup of a population based cohort in Sweden. J Urol. 2006;175:1337-40.

Andriole GL, Crawford ED, Grubb RL 3rd, Buys SS, Chia D, Church TR, Fouad MN, Gelmann EP, Kvale PA, Reding DJ, Weissfeld JL, Yokochi LA, O'Brien B, Clapp JD, Rathmell JM, Riley TL, Hayes RB, Kramer BS, Izmirlian G, Miller AB, Pinsky PF, Prorok PC, Gohagan JK, Berg CD; PLCO Project Team. Mortality results from a randomized prostate-cancer screening trial. N Engl J Med. 2009;360:1310-9.

Andriole GL, Bostwick DG, Brawley OW, Gomella LG, Marberger M, Montorsi F, Pettaway CA, Tammela TL, Teloken C, Tindall DJ, Somerville MC, Wilson TH, Fowler IL, Rittmaster RS; REDUCE Study Group. Effect of dutasteride on the risk of prostate cancer. N Engl J Med. 2010;362:1192-202.

Andriole GL, Crawford ED, Grubb RL 3rd, Buys SS, Chia D, Church TR, Fouad MN, Isaacs C, Kvale PA, Reding DJ, Weissfeld JL, Yokochi LA, O'Brien B, Ragard LR, Clapp JD, Rathmell JM, Riley TL, Hsing AW, Izmirlian G, Pinsky PF, Kramer BS, Miller AB, Gohagan JK, Prorok PC; PLCO Project Team. 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 2012;104:125-32.

Aus G, Bergdahl S, Lodding P, Lilja H, Hugosson J. Prostate cancer screening decreases the absolute risk of being diagnosed with advanced prostate cancer--results from a prospective, population-based randomized controlled trial. Eur Urol. 2007;51:659-64.

Auvinen A, Raitanen J, Moss S, de Koning HJ, Hugosson J, Tammela T, Roobol M, Lilja H, Hakama M. Test sensitivity in the European prostate cancer screening trial: results from Finland, Sweden, and the Netherlands. Cancer Epidemiol Biomarkers Prev. 2009;18:2000-5.

Bach PB, Vickers AJ. Do the data support the comorbidity hypothesis for the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial results? J Clin Oncol. 2011;29:e387.

Baker SG, Lichtenstein P, Kaprio J, Holm N. Genetic susceptibility to prostate, breast, and colorectal cancer among Nordic twins. Biometrics 2005;61:55-63.

Bangma CH, van Schaik RH, Blijenberg BG, Roobol MJ, Lilja H, Stenman UH. On the use of prostate-specific antigen for screening of prostate cancer in European Randomised Study for Screening of Prostate Cancer. Eur J Cancer. 2010;46:3109-19.

Bergdahl AG, Aus G, Lilja H, Hugosson J. Risk of dying from prostate cancer in men randomized to screening: differences between attendees and nonattendees. Cancer. 2009;115:5672-9.

Bhopal R. Concepts of epidemiology. Integrating the ideas, theories, principles and methods of epidemiology, 2nd edition. Oxford University Press 2008, Oxford, United Kingdom.

Bill-Axelson A, Holmberg L, Ruutu M, Garmo H, Stark JR, Busch C, Nordling S, Häggman M, Andersson SO, Bratell S, Spångberg A, Palmgren J, Steineck G, Adami HO, Johansson JE; SPCG-4 Investigators. Radical prostatectomy versus watchful waiting in early prostate cancer. N Engl J Med. 2011;364:1708-17.

Black WC, Haggstrom DA, Welch HG. All-cause mortality in randomized trials of cancer screening. J Natl Cancer Inst. 2002;94:167-73.

Bonovas S, Filioussi K, Sitaras NM. Statin use and the risk of prostate cancer: A meta-analysis of 6 randomized clinical trials and 13 observational studies. Int J Cancer. 2008;123:899-904.

Bostwick DG, Liu L, Brawer MK, Qian J. High-grade prostatic intraepithelial neoplasia. Rev Urol. 2004;6:171-9.

Bratt O, Kristoffersson U, Lundgren R, Olsson H. Familial and hereditary prostate cancer in southern Sweden. A population-based case-control study. Eur J Cancer. 1999;35:272-7.

Bratt O, Damber JE, Emanuelsson M, Grönberg H. Hereditary prostate cancer: clinical characteristics and survival. J Urol. 2002;167:2423-6.

Bray F, Lortet-Tieulent J, Ferlay J, Forman D, Auvinen A. Prostate cancer incidence and mortality trends in 37 European countries: an overview. Eur J Cancer. 2010;46:3040-52.

Breslow N, Chan CW, Dhom G, Drury RAB, Franks LM, Gellei B, Lee YS, Lundberg S, Sparke B, Sternbya NH, Tulinius H. Latent carcinoma of prostate at autopsy in seven areas. Collaborative study organized by the International Agency for Research on Cancer, Lyons, France. Int J Cancer. 1977;20:680–8.

Bruner DW, Moore D, Parlanti A, Dorgan J, Engstrom P. Relative risk of prostate cancer for men with affected relatives: systematic review and meta-analysis. Int J Cancer. 2003;107:797-803.

Bul M, van den Bergh RC, Zhu X, Rannikko A, Vasarainen H, Bangma CH, Schröder FH, Roobol MJ. Outcomes of initially expectantly managed patients with low or intermediate risk screen-detected localized prostate cancer. BJU Int. 2012 Aug 29. [Epub ahead of print]

Buschemeyer WC 3rd, Freedland SJ. Obesity and prostate cancer: epidemiology and clinical implications. Eur Urol. 2007;52:331-43.

Bushman W. Etiology, epidemiology, and natural history of benign prostatic hyperplasia. Urol Clin North Am. 2009;36:403-15.

Carlsson SV, Holmberg E, Moss SM, Roobol MJ, Schröder FH, Tammela TL, Aus G, Auvinen AP, Hugosson J. No excess mortality after prostate biopsy: results from the European Randomized Study of Screening for Prostate Cancer. BJU Int 2011;107:1912-7.

Carter BS, Beaty TH, Steinberg GD, Childs B, Walsh PC. Mendelian inheritance of familial prostate cancer. Proc Natl Acad Sci USA. 1992;89:3367-71.

Carter BS, Bova GS, Beaty TH, Steinberg GD, Childs B, Isaacs WB, Walsh PC. Hereditary prostate cancer: epidemiologic and clinical features. J Urol. 1993;150:797-802.

Catalona WJ, Smith DS, Ratliff TL, Dodds KM, Coplen DE, Yuan JJ, Petros JA, Andriole GL. Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. N Engl J Med. 1991;324:1156-61.

Catalona WJ, Richie JP, Ahmann FR, Hudson MA, Scardino PT, Flanigan RC, deKernion JB, Ratliff TL, Kavoussi LR, Dalkin BL, et al. Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men. J Urol. 1994;151:1283-90.

Cerhan JR, Parker AS, Putnam SD, Chiu BC, Lynch CF, Cohen MB, Torner JC, Cantor KP. Family history and prostate cancer risk in a population-based cohort of Iowa men. Cancer Epidemiol Biomarkers Prev. 1999;8:53-60.

Christensen GB, Baffoe-Bonnie AB, George A, Powell I, Bailey-Wilson JE, Carpten JD, Giles GG, Hopper JL, Severi G, English DR, Foulkes WD, Maehle L, Moller P, Eeles R, Easton D, Badzioch MD, Whittemore AS, Oakley-Girvan I, Hsieh CL, Dimitrov L, Xu J, Stanford JL, Johanneson B, Deutsch K, McIntosh L, Ostrander EA, Wiley KE, Isaacs SD, Walsh PC, Isaacs WB, Thibodeau SN, McDonnell SK, Hebbring S, Schaid DJ, Lange EM, Cooney KA, Tammela TL, Schleutker J, Paiss T, Maier C, Grönberg H, Wiklund F, Emanuelsson M, Farnham JM, Cannon-Albright LA, Camp NJ; International Consortium for Prostate Cancer Genetics. Genome-wide linkage analysis of 1,233 prostate cancer pedigrees from the International Consortium for Prostate Cancer Genetics using novel sumLINK and sumLOD analyses. Prostate. 2010;70:735-44.

Cohen YC, Liu KS, Heyden NL, Carides AD, Anderson KM, Daifotis AG, Gann PH. Detection bias due to the effect of finasteride on prostate volume: a modeling approach for analysis of the Prostate Cancer Prevention Trial. J Natl Cancer Inst 2007;99:1366–1374.

Collin SM, Martin RM, Metcalfe C, Gunnell D, Albertsen PC, Neal D, Hamdy F, Stephens P, Lane JA, Moore R, Donovan J. Prostate-cancer mortality in the USA and UK in 1975-2004: an ecological study. Lancet Oncol. 2008;9:445-52.

Collin SM, Metcalfe C, Refsum H, Lewis SJ, Zuccolo L, Smith GD, Chen L, Harris R, Davis M, Marsden G, Johnston C, Lane JA, Ebbing M, Bønaa KH, Nygård O, Ueland PM, Grau MV, Baron JA, Donovan JL, Neal DE, Hamdy FC, Smith AD, Martin RM. Circulating folate, vitamin B12, homocysteine, vitamin B12 transport proteins, and risk of prostate cancer: a case-control study, systematic review, and meta-analysis. Cancer Epidemiol Biomarkers Prev. 2010;19:1632-42.

Colloca G, Venturino A. The evolving role of familial history for prostate cancer. Acta Oncol. 2011;50:14-24.

Cook LS, Goldoft M, Schwartz SM, Weiss NS. Incidence of adenocarcinoma of the prostate in Asian immigrants to the United States and their descendants. J Urol. 1999;161:152-5.

Cooperberg MR, Carroll PR, Klotz L. Active surveillance for prostate cancer: progress and promise. J Clin Oncol. 2011;29:3669-76.

Crawford ED, Grubb R 3rd, Black A, Andriole GL Jr, Chen MH, Izmirlian G, Berg CD, D'Amico AV. Comorbidity and mortality results from a randomized prostate cancer screening trial. J Clin Oncol. 2011;29:355-61.

Cremers RG, Karim-Kos HE, Houterman S, Verhoeven RH, Schröder FH, van der Kwast TH, Kil PJ, Coebergh JW, Kiemeney LA. Prostate cancer: trends in incidence, survival and mortality in the Netherlands, 1989-2006. Eur J Cancer. 2010;46:2077-87.

Croswell JM, Kramer BS, Kreimer AR, Prorok PC, Xu JL, Baker SG, Fagerstrom R, Riley TL, Clapp JD, Berg CD, Gohagan JK, Andriole GL, Chia D, Church TR, Crawford ED, Fouad MN, Gelmann EP, Lamerato L, Reding DJ, Schoen RE. Cumulative incidence of false-positive results in repeated, multimodal cancer screening. Ann Fam Med. 2009;7:212-22.

Curado MP, Edwards B, Shin HR, Storm H, Ferlay J, Heanue M, Boyle P (editors). Cancer Incidence in Five Continents, Vol. IX. IARC Scientific Publications 2007; No. 160, Lyon, IARC.

Cuzick J, Edwards R, Segnan N. Adjusting for non-compliance and contamination in randomized clinical trials. Stat Med. 1997;16:1017-29.

Dennis LK. Meta-analysis for combining relative risks of alcohol consumption and prostate cancer. Prostate. 2000;42:56-66.

Dennis LK, Lynch CF, Torner JC. Epidemiologic association between prostatitis and prostate cancer. Urology. 2002;60:78-83.

Djulbegovic M, Beyth RJ, Neuberger MM, Stoffs TL, Vieweg J, Djulbegovic B, Dahm P. Screening for prostate cancer: systematic review and meta-analysis of randomised controlled trials. BMJ. 2010;341:c4543.

Draisma G, Etzioni R, Tsodikov A, Mariotto A, Wever E, Gulati R, Feuer E, de Koning H. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. J Natl Cancer Inst. 2009;101:374-83.

Eaton NE, Reeves GK, Appleby PN, Key TJ. Endogenous sex hormones and prostate cancer: a quantitative review of prospective studies. Br J Cancer. 1999;80:930-4.

Egevad L, Norlén BJ, Norberg M. The value of multiple core biopsies for predicting the Gleason score of prostate cancer. BJU Int. 2001;88:716-21.

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. 2006;175:1605-12.

Engholm G, Ferlay J, Christensen N, Gjerstorff ML, Johannesen TB, Klint Å, Køtlum JE, Ólafsdóttir E, Pukkala E and Storm HH. NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 4.0. Association of the Nordic Cancer Registries. Danish Cancer Society (http://www.ancr.nu), 2011. (Accessed October 14th, 2011)

Epstein JI, Herawi M. Prostate needle biopsies containing prostatic intraepithelial neoplasia or atypical foci suspicious for carcinoma: implications for patient care. J Urol 2006;175:820–834.

Epstein JI. Precursor lesions to prostatic adenocarcinoma. Virchows Arch. 2009;454:1-16.

Epstein JI. An update of the Gleason grading system. J Urol. 2010;183:433-40.

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. 1999;91:1033-9.

Etzioni R, Penson DF, Legler JM, di Tommaso D, Boer R, Gann PH, Feuer EJ. Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends. J Natl Cancer Inst. 2002;94:981-90.

Etzioni R, Tsodikov A, Mariotto A, Szabo A, Falcon S, Wegelin J, DiTommaso D, Karnofski K, Gulati R, Penson DF, Feuer E. Quantifying the role of PSA screening in the US prostate cancer mortality decline. Cancer Causes Control. 2008;19:175-81.

European Association of Urology Guidelines on Prostate Cancer, 2011. www.uroweb.org. (Accessed September 26th, 2011).

Feinleib M, Zelen M. Some pitfalls in the evaluation of screening programs. Arch Environ Health. 1969;19:412-5.

Ferlay J, Shin HR, Bray F, Forman D, Mathers C and Parkin DM. GLOBOCAN 2008 v1.2, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 [Internet]. Lyon, France: International Agency for Research on Cancer; 2010. Available from: http://globocan.iarc.fr. (Accessed September 30th, 2011).

Ficarra V, Sooriakumaran P, Novara G, Schatloff O, Briganti A, Van der Poel H, Montorsi F, Patel V, Tewari A, Mottrie A. Systematic review of methods for reporting combined outcomes after radical prostatectomy and proposal of a novel system: the survival, continence, and potency (SCP) classification. Eur Urol 2012;61:541-8.

Finne P, Auvinen A, Maattanen L, Tammela TL, Ruutu M, Juusela H, Martikainen P, Hakama M, Stenman UH. Diagnostic value of free prostate-specific antigen among men with a prostate-specific antigen level of <3.0 microg per liter. European Urology. 2008;54:362-70.

Finne P, Fallah M, Hakama M, Ciatto S, Hugosson J, de Koning H, Moss S, Nelen V, Auvinen A. Lead-time in the European Randomised Study of Screening for Prostate Cancer. Eur J Cancer. 2010;46:3102-8.

Ford ME, Havstad SL, Demers R, Cole Johnson C. Effects of false positive prostate cancer screening results on subsequent prostate cancer screening behavior. Cancer Epidemiol Biomarkers Prev 2005;14:190-4.

Fowler FJ Jr, Barry MJ, Walker-Corkery B, Caubet JF, Bates DW, Lee JM, Hauser A, McNaughton-Collins M. The impact of a suspicious prostate biopsy on patients' psychological, socio-behavioral, and medical care outcomes. J Gen Intern Med. 2006;21:715-21.

Gong Z, Neuhouser ML, Goodman PJ, Albanes D, Chi C, Hsing AW, Lippman SM, Platz EA, Pollak MN, Thompson IM, Kristal AR. Obesity, diabetes, and risk of prostate cancer: results from the prostate cancer prevention trial. Cancer Epidemiol Biomarkers Prev. 2006;15:1977-83.

Gosselaar C, Kranse R, Roobol MJ, Roemeling S, Schröder FH. The interobserver variability of digital rectal examination in a large randomized trial for the screening of prostate cancer. Prostate. 2008;68:985-93.

Greene KL, Albertsen PC, Babaian RJ, Carter HB, Gann PH, Han M, Kuban DA, Sartor AO, Stanford JL, Zietman A, Carroll P. Prostate specific antigen best practice statement: 2009 update. J Urol. 2009;182:2232-41.

Grubb RL 3rd, Pinsky PF, Greenlee RT, Izmirlian G, Miller AB, Hickey TP, Riley TL, Mabie JE, Levin DL, Chia D, Kramer BS, Reding DJ, Church TR, Yokochi LA, Kvale PA, Weissfeld JL, Urban DA, Buys SS, Gelmann EP, Ragard LR, Crawford ED, Prorok PC, Gohagan JK, Berg CD, Andriole GL. 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. 2008;102:1524-30.

Gupta D, Lammersfeld CA, Trukova K, Lis CG. Vitamin D and prostate cancer risk: a review of the epidemiological literature. Prostate Cancer Prostatic Dis. 2009;12:215-26.

Hakama M, Auvinen A, Day NE, Miller AB. Sensitivity in cancer screening. J Med Screen. 2007;14:174-7.

Hakama M, Auvinen A. Cancer Screening. In: International Encyclopedia of Public Health, Heggenhougen HK, Quah SR (eds.). pp. 464-480. 1st edition. Academic Press, United States, 2008.

Hara M, Koyanagi Y, Inoue T, Fukuyama T. Some physico-chemical characteristics of "gamma-seminoprotein", an antigenic component specific for human seminal plasma. Forensic immunological study of body fluids and secretion [Japanese]. Nihon Hoigaku Zasshi 1971;25:322–4.

Hayes RB, Liff JM, Pottern LM, Greenberg RS, Schoenberg JB, Schwartz AG, Swanson GM, Silverman DT, Brown LM, Hoover RN, et al. Prostate cancer risk in U.S. blacks and whites with a family history of cancer. Int J Cancer. 1995;60:361-4.

Heesakkers RA, Hövels AM, Jager GJ, van den Bosch HC, Witjes JA, Raat HP, Severens JL, Adang EM, van der Kaa CH, Fütterer JJ, Barentsz J. MRI with a lymph-node-specific contrast agent as an alternative to CT scan and lymph-node dissection in patients with prostate cancer: a prospective multicohort study. Lancet Oncol. 2008;9:850-6.

Heidenreich A, Bellmunt J, Bolla M, Joniau S, Mason M, Matveev V, Mottet N, Schmid HP, van der Kwast T, Wiegel T, Zattoni F; European Association of Urology. EAU guidelines on prostate cancer. Part 1: screening, diagnosis, and treatment of clinically localised disease. Eur Urol. 2011;59:61-71.

Heijmink SW, Fütterer JJ, Strum SS, Oyen WJ, Frauscher F, Witjes JA, Barentsz JO. State-of-the-art uroradiologic imaging in the diagnosis of prostate cancer. Acta Oncol. 2011;50 Suppl 1:25-38.

Heijnsdijk EA, Wever EM, Auvinen A, Hugosson J, Ciatto S, Nelen V, Kwiatkowski M, Villers A, Páez A, Moss SM, Zappa M, Tammela TL, Mäkinen T, Carlsson S, Korfage IJ, Essink-Bot ML, Otto SJ, Draisma G, Bangma CH, Roobol MJ, Schröder FH, de Koning HJ. Quality-of-life effects of prostate-specific antigen screening. N Engl J Med. 2012;367:595-605.

Helpap B, Egevad L. The significance of modified Gleason grading of prostatic carcinoma in biopsy and radical prostatectomy specimens. Virchows Arch. 2006;449:622-7.

Hickey K, Do KA, Green A. Smoking and prostate cancer. Epidemiol Rev. 2001;23:115-25.

Hodge KK, McNeal JE, Terris MK, Stamey TA. Random systematic versus directed ultrasound guided transrectal core biopsies of the prostate. J Urol. 1989;142:71-5.

Horner MJ, Ries LAG, Krapcho M, Neyman N, Aminou R, Howlader N, Altekruse SF, Feuer EJ, Huang L, Mariotto A, Miller BA, Lewis DR, Eisner MP, Stinchcomb DG, Edwards BK (eds). SEER Cancer Statistics Review, 1975-2006, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2006/, based on November 2008 SEER data submission, posted to the SEER web site, 2009. (Accessed September 28th, 2011).

Huang WY, Hayes R, Pfeiffer R, Viscidi RP, Lee FK, Wang YF, Reding D, Whitby D, Papp JR, Rabkin CS. Sexually transmissible infections and prostate cancer risk. Cancer Epidemiol Biomarkers Prev. 2008;17:2374-81.

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. 2004;100:1397-405.

Hugosson J, Carlsson S, Aus G, Bergdahl S, Khatami A, Lodding P, Pihl CG, Stranne J, Holmberg E, Lilja H. Mortality results from the Göteborg randomised population-based prostate-cancer screening trial. Lancet Oncol. 2010;11:725-32.

Hutchison GB, Shapiro S. Lead time gained by diagnostic screening for breast cancer. J Natl Cancer Inst. 1968;41:665-81.

Iczkowski KA, Lucia MS. Current perspectives on Gleason grading of prostate cancer. Curr Urol Rep. 2011;12:216-22.

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

International Union Against Cancer (UICC). TNM Classification of Malignant Tumours, 7th edition. Sobin LH, Gospodarowicz MK, Wittekind C (editors). Oxford: Wiley-Blackwell; 2009.

Ioannidis JP, Castaldi P, Evangelou E. A compendium of genome-wide associations for cancer: critical synopsis and reappraisal. J Natl Cancer Institute. 2010;102:846-58.

Jafari S, Etminan M, Afshar K. Nonsteroidal anti-inflammatory drugs and prostate cancer: a systematic review of the literature and meta-analysis. Can Urol Assoc J. 2009;3:323–30.

Jager GJ, Severens JL, Thornbury JR, de La Rosette JJ, Ruijs SH, Barentsz JO. Prostate cancer staging: should MR imaging be used? -- A decision analytic approach. Radiology. 2000;215:445-51.

Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiol Biomarkers Prev. 2010;19:1893-907.

Johns LE, Houlston RS. A systematic review and meta-analysis of familial prostate cancer risk. BJU Int. 2003;91:789-94.

Kalish LA, McDougal WS, McKinlay JB. Family history and the risk of prostate cancer. Urology. 2000;56:803-6.

Kaplan SA, Roehrborn CG, Meehan AG, Liu KS, Carides AD, Binkowitz BS, Heyden NL, Vaughan ED Jr. PCPT: evidence that finasteride reduces risk of most frequently detected intermediate-and high-grade (Gleason score 6 and 7) cancer. Urology 2009;73:935–939.

Klein EA, Thompson IM Jr, Tangen CM, Crowley JJ, Lucia MS, Goodman PJ, Minasian LM, Ford LG, Parnes HL, Gaziano JM, Karp DD, Lieber MM, Walther PJ, Klotz L, Parsons JK, Chin JL, Darke AK, Lippman SM, Goodman GE, Meyskens FL Jr, Baker LH. Vitamin E and the risk of prostate cancer: the Selenium and Vitamin E Cancer Prevention Trial (SELECT). JAMA. 2011;306:1549-56.

Kolonel, L.N. Fat, meat, and prostate cancer. Epidemiol. Rev. 2001;23:72–81.

de Koning HJ, Liem MK, Baan CA, Boer R, Schröder FH, Alexander FE; ERSPC. Prostate cancer mortality reduction by screening: power and time frame with complete enrollment in the European Randomised Screening for Prostate Cancer (ERSPC) trial. Int J Cancer. 2002;98:268-73.

de Koning HJ, Blom J, Merkelbach JW, Raaijmakers R, Verhaegen H, Van Vliet P, Nelen V, Coebergh JW, Hermans A, Ciatto S, Mäkinen T. Determining the cause of death in randomized screening trial(s) for prostate cancer. BJU Int. 2003;92 Suppl 2:71-8.

Koskimäki J, Hakama M, Huhtala H, Tammela TL. Prevalence of lower urinary tract symptoms in Finnish men: a population-based study. Br J Urol. 1998;81:364-9.

van der Kwast TH, Ciatto S, Martikainen PM, Hoedemaeker R, Laurila M, Pihl CG, Hugosson J, Neetens I, Nelen V, Di Lollo S, Roobol MJ, Määtänen L, Santonja C, Moss S, Schröder FH. Detection rates of high-grade prostate cancer during subsequent screening visits. Results of the European Randomized Screening Study for Prostate Cancer. Int J Cancer. 2006;118:2538-42.

Käypä hoito: Eturauhassyöpä, 2007. Suomalaisen lääkäriseuran Duodecimin ja Suomen Urologiyhdistyksen asettama työryhmä. http://www.kaypahoito.fi/web/kh/suositukset/naytaartikkeli/tunnus/hoi11060. (Accessed September 26th, 2011).

Labrie F, Candas B, Dupont A, Cusan L, Gomez JL, Suburu RE, Diamond P, Lévesque J, Belanger A. Screening decreases prostate cancer death: first analysis of the 1988 Quebec prospective randomized controlled trial. Prostate. 1999;38:83-91.

Labrie F, Candas B, Cusan L, Gomez JL, Bélanger A, Brousseau G, Chevrette E, Lévesque J. Screening decreases prostate cancer mortality: 11-year follow-up of the 1988 Quebec prospective randomized controlled trial. Prostate 2004;59:311-8.

Lafata JE, Simpkins J, Lamerato L, Poisson L, Divine G, Johnson CC. The economic impact of false-positive cancer screens. Cancer Epidemiol Biomarkers Prev. 2004;13:2126-32.

Lahti RA, Penttilä A. The validity of death certificates: routine validation of death certification and its effects on mortality statistics. Forensic Sci Int. 2001;115:15-32.

Laurila M, van der Kwast T, Bubendorf L, di Lollo S, Pihl CG, Ciatto S, Hugosson J, Määttänen L, Roobol MJ, Kujala PM. Detection rates of cancer, high grade PIN and atypical lesions suspicious for cancer in the European Randomized Study of Screening for Prostate Cancer. Eur J Cancer. 2010;46:3068-72.

van Leeuwen PJ, Roobol MJ, Kranse R, Zappa M, Carlsson S, Bul M, Zhu X, Bangma CH, Schröder FH, Hugosson J. Towards an optimal interval for prostate cancer screening. Eur Urol. 2012;61:171-6.

Leissner KH, Tisell LE. The weight of the human prostate. Scand J Urol Nephrol. 1979;13:137-42.

Leitzmann MF, Stampfer MJ, Wu K, Colditz GA, Willett WC, Giovannucci EL. Zinc supplement use and risk of prostate cancer. J Natl Cancer Inst 2003;95:1004-7.

Lesko SM, Rosenberg L, Shapiro S. Family history and prostate cancer risk. Am J Epidemiol. 1996;144:1041-7.

Lesko SM, Louik C, Vezina R, Rosenberg L, Shapiro S. Vasectomy and prostate cancer. J Urol. 1999;161:1848-52.

Lilja H, Ulmert D, Björk T, Becker C, Serio AM, Nilsson JA, Abrahamsson PA, Vickers AJ, Berglund G. Long-term prediction of prostate cancer up to 25 years before diagnosis of prostate cancer using prostate kallikreins measured at age 44 to 50 years. J Clin Oncol. 2007;25:431-6.

Lucia MS, Epstein JI, Goodman PJ, Darke AK, Reuter VE, Civantos F, Tangen CM, Parnes HL, Lippman SM, La Rosa FG, Kattan MW, Crawford ED, Ford LG, Coltman CA Jr, Thompson IM. Finasteride and high-grade prostate cancer in the Prostate Cancer Prevention Trial. J Natl Cancer Inst. 2007;99:1375-83.

Ma RW, Chapman K. A systematic review of the effect of diet in prostate cancer prevention and treatment. J Hum Nutr Diet. 2009;22:187-99.

MacInnis RJ, English DR. Body size and composition and prostate cancer risk: systematic review and meta-regression analysis. Cancer Causes Control. 2006;17:989-1003.

MacLennan GT, Eisenberg R, Fleshman RL, Taylor JM, Fu P, Resnick MI, Gupta S. The influence of chronic inflammation in prostatic carcinogenesis: a 5-year followup study. J Urol. 2006;176:1012-6.

Mahmud SM, Franco EL, Aprikian AG. Use of non-steroidal anti-inflammatory drugs (NSAIDs) and prostate cancer risk: A meta-analysis. Int J Cancer. 2010;127:1680–91.

Malila N, Palva T, Malminiemi O, Paimela H, Anttila A, Hakulinen T, Järvinen H, Kotisaari ML, Pikkarainen P, Rautalahti M, Sankila R, Vertio H, Hakama M. Coverage and performance of colorectal cancer screening with the faecal occult blood test in Finland. J Med Screen. 2011;18:18-23.

Malmi H, Ruutu M, Määttänen L, Stenman UH, Juusela H, Tammela TL, Auvinen A. Why do men opt out of prostate-cancer screening? Attitudes and perception among participants and non-participants of a screening trial. BJU Int. 2010;106:472-7.

de Marzo AM, DeWeese TL, Platz EA, Meeker AK, Nakayama M, Epstein JI, Isaacs WB, Nelson WG. Pathological and molecular mechanisms of prostate carcinogenesis: implications for diagnosis, detection, prevention, and treatment. J Cell Biochem. 2004;91:459-77.

Matikainen MP, Pukkala E, Schleutker J, Tammela TL, Koivisto P, Sankila R, Kallioniemi OP. Relatives of prostate cancer patients have an increased risk of prostate and stomach cancers: a population-based, cancer registry study in Finland. Cancer Causes Control. 2001;12:223-30.

Meyer F, Têtu B, Bairati I, Lacombe L, Fradet Y. Prostatic intraepithelial neoplasia in TURP specimens and subsequent prostate cancer. Can J Urol. 2006;13:3255-60.

Miller AB, Yurgalevitch S, Weissfeld JL; Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial Project Team. Death review process in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. Control Clin Trials. 2000;21(6 Suppl):400S-406S.

Miller DC, Hafez KS, Stewart A, Montie JE, Wei JT. Prostate carcinoma presentation, diagnosis, and staging: an update form the National Cancer Data Base. Cancer. 2003;98:1169-78.

Montironi R, Mazzuccheli R, Scarpelli M, Lopez-Beltran A, Fellegara G, Algaba F. Gleason grading of prostate cancer in needle biopsies or radical prostatectomy specimens: contemporary approach, current clinical significance and sources of pathology discrepancies. BJU Int. 2005;95:1146-52.

Mordukhovich I, Reiter PL, Backes DM, Family L, McCullough LE, O'Brien KM, Razzaghi H, Olshan AF. A review of African American-white differences in risk factors for cancer: prostate cancer. Cancer Causes Control. 2011;22:341-57.

Morganti G, Gianferrari L, Cresseri A, Arrigoni G, Lovati G. Clinicostatistical and genetic research on neoplasms of the prostate. Acta Genet Stat Med. 1956-1957;6:304–5.

Mostofi FK. Histological typing of prostate tumours. International histological classification of tumors. WHO Geneva 1980; no. 22.

Moyer VA; on behalf of the U.S. Preventive Services Task Force. Screening for Prostate Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med 2012; May 21. [Epub ahead of print.]

Murad AS, Down L, Davey Smith G, Donovan JL, Athene Lane J, Hamdy FC, Neal DE, Martin RM. Associations of aspirin, nonsteroidal anti-inflammatory drug and paracetamol use with PSA-detected prostate cancer: findings from a large, population-based, case-control study (the ProtecT study). Int J Cancer. 2011;128:1442-8.

Mäkinen T, Tammela TL, Hakama M, Stenman UH, Rannikko S, Aro J, Juusela H, Määttänen L, Auvinen A. Prostate cancer screening within a prostate specific antigen range of 3 to 3.9 ng./ml.: a comparison of digital rectal examination and free prostate specific antigen as supplemental screening tests. J Urol. 2001;166:1339-42.

Mäkinen T, Auvinen A, Hakama M, Stenman UH, Tammela TL. Acceptability and complications of prostate biopsy in population-based PSA screening versus routine clinical practice: a prospective, controlled study. Urology. 2002;60:846-50.

Mäkinen T, Karhunen P, Aro J, Lahtela J, Määttänen L, Auvinen A. Assessment of causes of death in a prostate cancer screening trial. Int J Cancer. 2008;122:413-7.

Määttänen L, Auvinen A, Stenman UH, Rannikko S, Tammela T, Aro J, Juusela H, Hakama M. European randomized study of prostate cancer screening: first-year results of the Finnish trial. Br J Cancer. 1999;79:1210-4.

Määttänen L, Hakama M, Tammela TL, Ruutu M, Ala-Opas M, Juusela H, Martikainen P, Stenman UH, Auvinen A. Specificity of serum prostate-specific antigen determination in the Finnish prostate cancer screening trial. Br J Cancer. 2007;96:56-60.

National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365:395-409.

Nelen V, Thys G, Hermans A, D'Hooge K, Dourcy-Belle-Rose B, Coebergh JW, Roobol M, Denis L. Interval cancers in the Antwerp European randomised study of screening for prostate cancer study, using a 6 year screening interval. Eur J Cancer. 2010;46:3090-4.

Nelson W. Theory and applications of hazard plotting for censored failure data. Technometrics 1972;14:945-66.

Norberg M, Egevad L, Holmberg L, Sparén P, Norlén BJ, Busch C. The sextant protocol for ultrasound-guided core biopsies of the prostate underestimates the presence of cancer. Urology. 1997;50:562-6.

Norrish AE, McRae CU, Cohen RJ, Jackson RT. A population-based study of clinical and pathological prognostic characteristics of men with familial and sporadic prostate cancer. BJU Int. 1999;84:311-5.

Oesterling JE, Jacobsen SJ, Chute CG, Guess HA, Girman CJ, Panser LA, Lieber MM. Serum prostate-specific antigen in a community-based population of healthy men. Establishment of age-specific reference ranges. JAMA. 1993;270:860-4.

Oliver SE, May MT, Gunnell D. International trends in prostate-cancer mortality in the "PSA ERA". Int J Cancer. 2001;92:893-8.

Ostrander EA, Stanford JL. Genetics of prostate cancer: too many loci, too few genes. Am J Hum Genet. 2000;67:1367-75.

Otto SJ, van der Cruijsen IW, Liem MK, Korfage IJ, Lous JJ, Schröder FH, de Koning HJ. Effective PSA contamination in the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer. Int J Cancer. 2003;105:394-9.

Otto SJ, van Leeuwen PJ, Hoekstra JW, Merckelbach JW, Blom JH, Schröder FH, Roobol MJ, de Koning HJ. Blinded and uniform causes of death verification in cancer screening: a major influence on the outcome of a prostate cancer screening trial? Eur J Cancer. 2010;46:3061-7.

Palapattu GS, Sutcliffe S, Bastian PJ, Platz EA, De Marzo AM, Isaacs WB, Nelson WG. Prostate carcinogenesis and inflammation: emerging insights. Carcinogenesis. 2005;26:1170-81.

Parker PM, Rice KR, Sterbis JR, Chen Y, Cullen J, McLeod DG, Brassell SA. Prostate cancer in men less than the age of 50: a comparison of race and outcomes. Urology. 2011;78:110-5.

Partin AW, Mangold LA, Lamm DM, Walsh PC, Epstein JI, Pearson JD. Contemporary update of prostate cancer staging nomograms (Partin Tables) for the new millennium. Urology. 2001;58:843-8.

Pinsky P, Parnes H, Ford L. Estimating rates of true high-grade disease in the prostate cancer prevention trial. Cancer Prev Res (Phila) 2008;1:182–6.

Pogodin-Hannolainen D, Juusela H, Tammela TL, Ruutu M, Aro J, Määttänen L, Auvinen A. Prostate cancer screening: a survey of attitudes and practices among Finnish physicians in 1999 and 2007. J Med Screen. 2011;18:46-9.

Postma R, Roobol M, Schröder FH, van der Kwast TH. Lesions predictive for prostate cancer in a screened population: first and second screening round findings. Prostate. 2004;61:260-6.

Postma R, Schröder FH, van der Kwast TH. Atrophy in prostate needle biopsy cores and its relationship to prostate cancer incidence in screened men. Urology. 2005;65:745-9.

Postma R, Schröder FH, van Leenders GJ, Hoedemaeker RF, Vis AN, Roobol MJ, van der Kwast TH. Cancer detection and cancer characteristics in the European Randomized Study of Screening for Prostate Cancer (ERSPC)--Section Rotterdam. A comparison of two rounds of screening. Eur Urol. 2007;52:89-97.

Putzi MJ, De Marzo AM. Morphologic transitions between proliferative inflammatory atrophy and high-grade prostatic intraepithelial neoplasia. Urology. 2000;56:828-32.

Raaijmakers R, de Vries SH, Blijenberg BG, Wildhagen MF, Postma R, Bangma CH, Darte C, Schröder FH. hK2 and free PSA, a prognostic combination in predicting minimal prostate cancer in screen-detected men within the PSA range 4-10 ng/ml. Eur Urol. 2007;52:1358-64.

Redman MW, Tangen CM, Goodman PJ, Lucia MS, Coltman CA Jr, Thompson IM. Finasteride does not increase the risk of high-grade prostate cancer: a bias-adjusted modeling approach. Cancer Prev Res (Phila) 2008;1:174–81.

Rhodes T, Girman CJ, Jacobsen SJ, Roberts RO, Guess HA, Lieber MM. Longitudinal prostate growth rates during 5 years in randomly selected community men 40 to 79 years old. J Urol. 1999;161:1174-9.

Rittmaster RS, Fleshner NE, Thompson IM. Pharmacological approaches to reducing the risk of prostate cancer. Eur Urol. 2009;55:1064-73.

Roddam AW, Duffy MJ, Hamdy FC, Ward AM, Patnick J, Price CP, Rimmer J, Sturgeon C, White P, Allen NE; NHS Prostate Cancer Risk Management Programme. Use of prostate-specific antigen (PSA) isoforms for the detection of prostate cancer in men with a PSA level of 2-10 ng/ml: systematic review and meta-analysis. Eur Urol. 2005;48:386-99.

Roddam AW, Allen NE, Appleby P, Key TJ, Endogenous Hormones and Prostate Cancer Collaborative Group. Endogenous sex hormones and prostate cancer: a collaborative analysis of 18 prospective studies. J Natl Cancer Inst. 2008;100:170-83.

Rodríguez C, Calle EE, Miracle-McMahill HL, Tatham LM, Wingo PA, Thun MJ, Heath CW Jr. Family history and risk of fatal prostate cancer. Epidemiology. 1997;8:653-7.

Rohrmann S, Linseisen J, Key TJ, Jensen MK, Overvad K, Johnsen NF, Tjønneland A, Kaaks R, Bergmann MM, Weikert C, Naska A, Trichopoulou A, Trichopoulos D, Pala V, Sacerdote C, Palli D, Tumino R, Bueno-de-Mesquita HB, Vrieling A, González CA, Larrañaga N, Navarro C, Barricarte A, Quiros JR, Martínez-García C, Hallmans G, Stattin P, Manjer J, Wirfält E, Bingham S, Khaw KT, Egevad L, Ferrari P, Jenab M, Riboli E. Alcohol consumption and the risk for prostate cancer in the European Prospective Investigation into Cancer and Nutrition. Cancer Epidemiol Biomarkers Prev. 2008;17:1282-7.

Rommel FM, Agusta VE, Breslin JA, Huffnagle HW, Pohl CE, Sieber PR, Stahl CA. The use of prostate specific antigen and prostate specific antigen density in the diagnosis of prostate cancer in a community based urology practice. J Urol. 1994;151:88-93.

Roobol MJ, Grenabo A, Schröder FH, Hugosson J. Interval cancers in prostate cancer screening: comparing 2- and 4-year screening intervals in the European Randomized Study of Screening for Prostate Cancer, Gothenburg and Rotterdam. J Natl Cancer Inst. 2007;99:1296-303.

Roobol MJ. Contemporary role of prostate cancer gene 3 in the management of prostate cancer. Curr Opin Urol. 2011;21:225-9.

Rosario DJ, Lane JA, Metcalfe C, Donovan JL, Doble A, Goodwin L, Davis M, Catto JW, Avery K, Neal DE, Hamdy FC. Short term outcomes of prostate biopsy in men tested for cancer by prostate specific antigen: prospective evaluation within ProtecT study. BMJ 2012;344:d7894.

Rothwell PM, Fowkes FG, Belch JF, Ogawa H, Warlow CP, Meade TW. Effect of daily aspirin on long-term risk of death due to cancer: analysis of individual patient data from randomised trials. Lancet 2011;377:31-41.

Sakr WA, Grignon DJ, Crissman JD, Heilbrun LK, Cassin BJ, Pontes JJ, Haas GP. High grade prostatic intraepithelial neoplasia (HGPIN) and prostatic adenocarcinoma between the ages of 20-69: an autopsy study of 249 cases. In Vivo. 1994;8:439-43.

Sandblom G, Varenhorst E, Löfman O, Rosell J, Carlsson P. Clinical consequences of screening for prostate cancer: 15 years follow-up of a randomised controlled trial in Sweden. Eur Urol. 2004;46:717-24.

Scattoni V, Maccagnano C, Zanni G, Angiolilli D, Raber M, Roscigno M, Rigatti P, Montorsi F. Is extended and saturation biopsy necessary? Int J Urol. 2010;17:432-47.

Schlesinger C, Bostwick DG, Iczkowski KA. High-grade prostatic intraepithelial neoplasia and atypical small acinar proliferation: predictive value for cancer in current practice. Am J Surg Pathol. 2005;29:1201-7.

Schröder FH, van der Maas P, Beemsterboer P, Kruger AB, Hoedemaeker R, Rietbergen J, Kranse R. Evaluation of the digital rectal examination as a screening test for prostate cancer. Rotterdam section of the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst. 1998;90:1817-23.

Schröder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, Kwiatkowski M, Lujan M, Lilja H, Zappa M, Denis LJ, Recker F, Berenguer A, Määttänen L, Bangma CH, Aus G, Villers A, Rebillard X, van der Kwast T, Blijenberg BG, Moss SM, de Koning HJ, Auvinen A; ERSPC Investigators. Screening and prostate-cancer mortality in a randomized European study. N Engl J Med. 2009;360:1320-8.

Schröder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, Kwiatkowski M, Lujan M, Lilja H, Zappa M, Denis LJ, Recker F, Páez A, Määttänen L, Bangma CH, Aus G, Carlsson S, Villers A, Rebillard X, van der Kwast T, Kujala PM, Blijenberg BG, Stenman UH, Huber A, Taari K, Hakama M, Moss SM, de Koning HJ, Auvinen A; ERSPC Investigators. Prostate-cancer mortality at 11 years of follow-up. N Engl J Med 2012;366:981-90.

Shiels MS, Pfeiffer RM, Gail MH, Hall HI, Li J, Chaturvedi AK, Bhatia K, Uldrick TS, Yarchoan R, Goedert JJ, Engels EA. Cancer burden in the HIV-infected population in the United States. J Natl Cancer Inst. 2011;103:753-62.

Shimizu H, Ross RK, Bernstein L, Yatani R, Henderson BE, Mack TM. Cancers of the prostate and breast among Japanese and white immigrants in Los Angeles County. Br J Cancer. 1991;63:963-6.

Sirovich BE, Schwartz LM, Woloshin S. Screening men for prostate and colorectal cancer in the United States: does practice reflect the evidence? JAMA. 2003;289:1414-20.

Soos G, Tsakiris I, Szanto J, Turzo C, Haas PG, Dezso B. The prevalence of prostate carcinoma and its precursor in Hungary: an autopsy study. Eur Urol. 2005;48:739-44.

Stamey TA, Yang N, Hay AR, McNeal JE, Freiha FS, Redwine E. Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. N Engl J Med. 1987;317:909-16.

Stanford JL, Wicklund KG, McKnight B, Daling JR, Brawer MK. Vasectomy and risk of prostate cancer. Cancer Epidemiol Biomarkers Prev. 1999;8:881-6.

Stark JR, Perner S, Stampfer MJ, Sinnott JA, Finn S, Eisenstein AS, Ma J, Fiorentino M, Kurth T, Loda M, Giovannucci EL, Rubin MA, Mucci LA. Gleason score and lethal prostate cancer: does 3 + 4 = 4 + 3? J Clin Oncol. 2009;27:3459-64.

Stattin P, Holmberg E, Johansson JE, Holmberg L, Adolfsson J, Hugosson J; National Prostate Cancer Register (NPCR) of Sweden. Outcomes in localized prostate cancer: National Prostate Cancer Register of Sweden follow-up study. J Natl Cancer Inst. 2010;102:950-8.

Stenman UH, Leinonen J, Alfthan H, Rannikko S, Tuhkanen K, Alfthan O. A complex between prostate-specific antigen and alpha 1-antichymotrypsin is the major form of prostate-specific antigen in serum of patients with prostatic cancer: assay of the complex improves clinical sensitivity for cancer. Cancer Res. 1991;51:222-6.

Stephenson AJ, Abouassaly R, Klein EA. Chemoprevention of prostate cancer. Urol Clin North Am. 2010;37:11-21.

Strope SA, Andriole GL. Update on chemoprevention for prostate cancer. Curr Opin Urol. 2010;20:194-7.

Sutcliffe S, Giovannucci E, De Marzo AM, Leitzmann MF, Willett WC, Platz EA. Gonorrhea, syphilis, clinical prostatitis, and the risk of prostate cancer. Cancer Epidemiol Biomarkers Prev. 2006;15:2160-6.

Teppo L, Pukkala E, Lehtonen M. Data quality and quality control of a population based cancer registry. Experience in Finland. Acta Oncol 1994;33:365–9.

Terry PD, Rohan TE, Wolk A. Intakes of fish and marine fatty acids and the risks of cancers of the breast and prostate and of other hormone-related cancers: a review of the epidemiologic evidence. Am. J. Clin. Nutr. 2003;77:532-43.

Thompson IM, Goodman PJ, Tangen CM, Lucia MS, Miller GJ, Ford LG, Lieber MM, Cespedes RD, Atkins JN, Lippman SM, Carlin SM, Ryan A, Szczepanek CM, Crowley JJ, Coltman CA Jr. The influence of finasteride on the development of prostate cancer. N Engl J Med. 2003;349:215-24.

Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, Minasian LM, Ford LG, Lippman SM, Crawford ED, Crowley JJ, Coltman CA Jr. Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. N Engl J Med. 2004;350:2239-46.

Thompson IM, Ankerst DP, Chi C, Lucia MS, Goodman PJ, Crowley JJ, Parnes HL, Coltman CA Jr. Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower. JAMA. 2005;294:66-70.

Trabulsi EJ, Sackett D, Gomella LG, Halpern EJ. Enhanced transrectal ultrasound modalities in the diagnosis of prostate cancer. Urology. 2010;76:1025-33.

Ulmert D, O'Brien MF, Bjartell AS, Lilja H. Prostate kallikrein markers in diagnosis, risk stratification and prognosis. Nat Rev Urol. 2009;6:384-91.

Valeri A, Azzouzi R, Drelon E, Delannoy A, Mangin P, Fournier G, Berthon P, Cussenot O. Early-onset hereditary prostate cancer is not associated with specific clinical and biological features. Prostate. 2000;45:66-71.

Vickers AJ, Cronin AM, Aus G, Pihl CG, Becker C, Pettersson K, Scardino PT, Hugosson J, Lilja H. A panel of kallikrein markers can reduce unnecessary biopsy for prostate cancer: data from the European Randomized Study of Prostate Cancer Screening in Göteborg, Sweden. BMC Med. 2008;6:19.

Vickers AJ, Savage C, O'Brien MF, Lilja H. Systematic review of pretreatment prostate-specific antigen velocity and doubling time as predictors for prostate cancer. J Clin Oncol. 2009;27:398-403.

Vis AN, Van Der Kwast TH. Prostatic intraepithelial neoplasia and putative precursor lesions of prostate cancer: a clinical perspective. BJU Int. 2001;88:147–57.

Wang MC, Papsidero LD, Kuriyama M, Valenzuela LA, Murphy GP, Chu TM. Prostate antigen: a new potential marker for prostatic cancer. Prostate. 1981;2:89-96.

Wang W, Bergh A, Damber JE. Morphological transition of proliferative inflammatory atrophy to high-grade intraepithelial neoplasia and cancer in human prostate. Prostate. 2009;69:1378-86.

Weir HK, Thun MJ, Hankey BF, Ries LA, Howe HL, Wingo PA, Jemal A, Ward E, Anderson RN, Edwards BK. Annual report to the nation on the status of cancer, 1975-2000, featuring the uses of surveillance data for cancer prevention and control. J Natl Cancer Inst. 2003;95:1276-99.

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

Whittemore AS, Wu AH, Kolonel LN, John EM, Gallagher RP, Howe GR, West DW, Teh CZ, Stamey T. Family history and prostate cancer risk in black, white, and Asian men in the United States and Canada. Am J Epidemiol. 1995;141:732-40.

Wilson JMG, Jungner G. Principles and practice of screening for disease. Geneva: WHO; 1968.

Wolk A. Diet, lifestyle and risk of prostate cancer. Acta Oncologica. 2005;44:277-81.

Wolters T, Roobol MJ, Schröder FH, van der Kwast TH, Roemeling S, van der Cruijsen-Koeter IW, Bangma CH, van Leenders GJ. Can non-malignant biopsy features identify men at increased risk of biopsy-detectable prostate cancer at re-screening after 4 years? BJU Int 2008;101:283-88.

Wolters T, Roobol MJ, Steyerberg EW, van den Bergh RC, Bangma CH, Hugosson J, Ciatto S, Kwiatkowski M, Villers A, Luján M, Nelen V, Tammela TL, Schröder FH. The effect of study arm on prostate cancer treatment in the large screening trial ERSPC. Int J Cancer. 2010;126:2387-93.

World Health Organization, mortality database 2011. http://www.who.int/whosis/mort/download/en/index.html. (Accessed October 3rd, 2011.)

Wright ME, Chang SC, Schatzkin A, Albanes D, Kipnis V, Mouw T, Hurwitz P, Hollenbeck A, Leitzmann MF. Prospective study of adiposity and weight change in relation to prostate cancer incidence and mortality. Cancer 2007;109:675–84.

Wu GH, Auvinen A, Yen AM, Hakama M, Tammela TL, Stenman UH, Kujala P, Ruutu M, Chen HH. The impact of interscreening interval and age on prostate cancer screening with prostate-specific antigen. Eur Urol. 2012;61:1011-8.

Zani EL, Clark OA, Rodrigues Netto N Jr. Antibiotic prophylaxis for transrectal prostate biopsy. Cochrane Database Syst Rev. 2011;(5):CD006576.

Zeegers MP, Jellema A, Ostrer H. Empiric risk of prostate carcinoma for relatives of patients with prostate carcinoma: a meta-analysis. Cancer. 2003;97:1894-903.

Zeliadt SB, Etzioni R, Ramsey SD, Penson DF, Potosky AL. Trends in treatment costs for localized prostate cancer: the healthy screenee effect. Med Care. 2007;45:154-9.

Zhu X, van Leeuwen PJ, Bul M, Bangma CH, Roobol MJ, Schröder FH. Identifying and characterizing "escapes"-men who develop metastases or die from prostate cancer despite screening (ERSPC, section Rotterdam). Int J Cancer. 2011;129:2847-54.

Zu K, Giovannucci E. Smoking and aggressive prostate cancer: a review of the epidemiologic evidence. Cancer Causes Control. 2009;20:1799-810.

Original communications

I

Kilpeläinen TP, Tammela TL, Määttänen L, Kujala P, Stenman UH, Ala-Opas M, Murtola TJ, Auvinen A. False-positive screening results in the Finnish prostate cancer screening trial. Br J Cancer. 2010;102:469-74.

False Positive Screening Results in the Finnish Prostate Cancer Screening Trial

Tuomas P Kilpeläinen* [1,2], Teuvo LJ Tammela [1], Liisa Määttänen [3], Paula Kujala [4], Ulf-Håkan Stenman [5], Martti Ala-Opas [6], Teemu J Murtola [2,7], Anssi Auvinen [2]

(1) University of Tampere and Tampere University Hospital, Department of Urology, Tampere, Finland
(2) Tampere School of Public Health, University of Tampere, Finland
(3) Finnish Cancer Registry, Helsinki, Finland
(4) Tampere University Hospital, Department of Pathology, Tampere, Finland
(5) Helsinki University Hospital, Department of Clinical Chemistry, Helsinki, Finland
(6) Helsinki University Hospital, Department of Urology, Helsinki, Finland
(7) Central Finland Central Hospital, Department of Surgery, Jyväskylä, Finland

***Address for correspondence:** Dr.* Tuomas P. Kilpeläinen (tuomas.kilpelainen@uta.fi), Tampere School of Public Health, University of Tampere, FIN-33014 Tampere, Finland. Fax +358335516057.

Keywords: Mass screening; Prostatic neoplasms; Prostate-specific antigen; Sensitivity and specificity; Randomized controlled trials

ABSTRACT: BACKGROUND: There is evidence that prostate cancer (PC) screening with prostate-specific antigen (PSA) serum test decreases PC mortality, but screening has adverse effects, such as high false positive (FP) rate. We investigated the proportion of FPs in a population-based randomized screening trial in Finland.

METHODS: Finland is the largest centre in the European Randomized Study of Screening for Prostate Cancer. We have completed three screening rounds with a 4-year screening interval (mean follow-up time 9.2 years) using a PSA cut-off level of 4.0 ng/ml; in addition men with PSA 3.0-3.9 and a positive auxiliary test were referred. A FP result was defined as a positive screening result without cancer in biopsy within 1 year from the screening test.

RESULTS: The proportion of FP screening results varied 3.3-12.1 % per round. Of the screened men, 12.5 % had at least one FP during three rounds. The risk of next-round PC following a FP result was 12.3-19.7 % vs. 1.4-3.7 % following a screen-negative result (depending on the screening round), RR 3.6-9.9. More than half of the men with one FP result had another one at subsequent screen. Men with a FP result were 1.5-2.0 times more likely to not participate in subsequent rounds compared to men with normal screening result (21.6-29.6 % vs. 14.0-16.7 %).

CONCLUSION: A FP result is a common adverse effect of PC screening and affects at least every eighth man screened repeatedly, even when using a relatively high cut-off level. FP men constitute a special group that receives unnecessary interventions but may harbour missed cancers. New strategies are needed for risk stratification in PC screening to minimize the proportion of FP men.

INTRODUCTION

Prostate cancer (PC) is the most common cancer in most industrialized countries [Parkin et al, 2005]. Its incidence increased steadily from the 1980's onward, as the increased use of transurethral resection of the prostate (TURP) for benign prostatic hyperplasia (BPH) resulted in more (incidental) PC diagnoses [Merrill et al, 1999]. A steep rise in the incidence of PC was observed in the 1990's when the prostate-specific antigen (PSA) test as a diagnostic tool was adapted widely [Welch et al, 2009]. Lately, the incidence of PC has been decreasing in many countries [Welch et al, 2009].

Screening for PC with PSA has become one of the most controversial public health issues. The two major screening trials in Europe and in the USA have provided inconsistent results concerning mortality effects of PSA-based PC screening [Andriole et al, 2009; Schröder et al, 2009]. While it is essential to determine mortality and quality of life effects of screening, it is also important to evaluate the sensitivity and specificity of the screening test to ensure best possible screening protocol, i.e. maximize the benefits and minimize the harms of screening. The proportion of false positive (FP) screening results indicates one aspect of adverse effects of screening, in addition to overdiagnosis and overtreatment.

The FP results are related to the specificity of the screening test. Specificity represents the ability of a test or test protocol to identify those free of the target disorder. Specificity is calculated as the ratio of the frequency of the true negative results (those with a negative test and without the target disease) to the sum of the frequencies of the true negatives and false positives (those with a positive test but free of the disease). Hence the proportion of FP results is 1-specificity.

FP screening results are common in PC screening, as PSA is an organ-specific, but not a disease-specific marker [Stenman et al, 2000]. It has been reported previously that approximately 70 % of men with elevated PSA do not have PC [Catalona et al, 1994; Schröder et al, 1998]. The proportion of FP results is likely to increase with age, as prostatic diseases, such as chronic prostatitis and BPH, become more common [Koskimäki et al, 1998; Rhodes et al, 1999; Wright et al, 2002]. The proportion of FP results has been estimated to be 7-8 % [Lafata et al, 2004; Määttänen et al, 2007] per screen (with one year of follow-up after the test). In repeated screening, the cumulative proportion of FP results was recently estimated as 10.4 % with four PSA tests and over three years of follow-up [Croswell et al, 2009].

The purpose of our study was to assess the proportion of FP results in a population-based randomized controlled trial in Finland during three screening rounds. We evaluated whether men with a FP result are at greater risk of decreased screening compliance, subsequent PC, or repeated FP result(s). We also investigated how many biopsies men with FP results undergo and whether the use of medication for BPH affects FP rates.

MATERIALS & METHODS

The Finnish Prostate Cancer Screening Study is the largest component of the European Randomized Study of Screening for Prostate Cancer (ERSPC), which is a multicenter randomized trial. The Finnish trial comprises 80,255 men born in 1929-1944 (aged 55, 59, 63 or 67 years at entry) and residing either in Helsinki or Tampere metropolitan area. Men with a previous PC diagnosis were excluded. The subjects were identified from the Finnish Population Registry. A random sample of 8,000 men was allocated to the screening arm annually in 1996-1999 and the remaining men formed the control group that received no interventions nor were they contacted. This analysis covered only the screened men.

The men in the screening arm were sent an invitation letter along with a brief overview of the trial, a questionnaire about urological symptoms, family history of PC, previous PSA tests and an informed consent form.

The men in the screening arm were invited to give a blood sample at a local cancer society clinic in Helsinki or Tampere. Men with PSA ≥ 4ng/ml were referred to a urological clinic for diagnostic examinations including digital rectal examination (DRE), transrectal ultrasound and biopsy. Initially a sextant biopsy was used, but this was increased to 10-12 biopsy cores in 2002. Men with PSA level of 3.0 – 3.9 ng/ml were referred to an additional test, which was DRE during 1996-1998 and since 1999 free/total PSA (F/T PSA) ratio with a cut-off point of 16 %. Men with a suspicious DRE or F/T PSA ratio <16 % were referred to diagnostic examinations similar to those with PSA ≥ 4.0 ng/ml.

All the laboratory analyses were carried out at the Department of Clinical Chemistry, Helsinki University Hospital. The serum concentrations of total PSA were analyzed by both Hybritech Tandem-E and Wallac Delfia assays. The free/total PSA ratio was determined with the Wallac ProStatus free/total PSA assay.

The men in the screening arm were then re-invited in a similar manner four and eight years after the first screen to the second and third screening rounds (though men older than 71 years of age were no longer invited because the core age-group in the protocol was 55-69 years of age). The first screening round was carried out in 1996-1999, the second in 2000-2003 and the third in 2004-2007. The common closing date of follow-up was December 31th, 2007 with mean follow-up of 9.2 years. All the men in the screening arm were invited to each round regardless of their participation in the previous round(s). Men diagnosed with PC were not re-invited, neither men who had emigrated from the study area or had died. Information on vital status and place of residence was obtained from the Population Register Centre.

Diagnosis of PC was based on histopathologic examination. A re-biopsy within 2 months was indicated if the primary histopathologic diagnosis was prostatic intraepithelial neoplasia, atypical small acinar proliferation or unconfirmed suspicion of PC, or if the PSA level was ≥ 10 ng/ml. The decision of re-biopsying a patient after a negative biopsy was made by the attending physician, who did not always comply with the protocol of the screening trial. Therefore, some re-biopsies were performed with less strict criteria and some postponed further than protocol defined time frames. The definition for a FP result was a positive screening result (based on both total PSA and either DRE or free/total PSA ratio) and consequent diagnostic work-up with no histopathologic diagnosis of PC in the biopsy within

one year from the PSA-test. The men who had a positive screening result but did not undergo biopsy according the screening protocol were not analyzed in this study.

Data on cancers detected outside the screening protocol were obtained from the nationwide, population-based Finnish Cancer Registry, which has 99 % coverage of all solid cancers diagnosed in Finland [Teppo et al, 1994]. Information on cancer incidence as well as vital status was available until the end of 2007. Data on BPH medication (finasteride or alpha-blockers) use at the time of screening (in 1996-2004) was obtained by linking the study population to the prescription drug database of the Social Insurance Institution of Finland (SII, www.kela.fi). SII is a governmental agency providing reimbursements to the Finnish citizens for the cost of drugs prescribed by physicians (with the exception of hospital inpatients).

The 95% confidence intervals (CI) for risks and proportions were calculated based on basic standard error formulas. Generalized linear model for the binomial distribution with a logarithmic link function was used to calculate risk ratios and their 95% CIs. The events were FP screening results and diagnosis of PC, with risk ratios indicating relative frequencies of outcomes in the groups to be compared. Statistical analyses were performed using Stata 8.2 (StataCorp, Texas, USA).

RESULTS

Out of 30,195 men in the screening arm, 23,771 (78.7 %) participated in at least one screening round, and 10,327 men (52.1 % of those invited to all rounds) participated in all the three rounds.

Altogether 1,611 cancers were detected by screening, of which 543 in the first round with a detection rate (DR) of 2.6 %, 613 (DR 3.3 %) in the second and 455 (DR 3.6 %) in the third round. The overall risk for a FP result was 6.4 % in the first, 8.0 % in the second and 7.8 % the third round. The risk of a FP result varied from 3.3 to 12.1 % per round, depending on screening round and age (Table 1). When men of similar age at screening were compared, the proportion of FP results was lower at repeat screening compared with the first round.

Of the men with a screen-positive result in the first round, 67.3 % turned out to be FP and 27.5 % were diagnosed with PC (5.2 % of screen-positive men were not biopsied according the protocol). In the second round, 64.6 % of the screen-positive findings were FP and 26.6 % PC, while in the third round 60.7 % were FP and 27.7 % PC. There was little variation by age (results not shown).

Of the 23,771 men who participated at least once during the three rounds, 12.5 % (CI 12.1-12.9) had at least one FP result**.** The proportion of men with at least one FP result during the screening programme increased consistently with age from 9.0 % in the youngest age cohort to 15.7 % in the oldest age cohort (with only two screening rounds). Of the 10,327 men who participated in all three rounds, 1,193 (11.6 %, CI 10.9-12.2 %) had at least one FP result. Of them, 1.2 % (CI 1.0-1.4 %) had a FP result in all three rounds, 2.8 % (CI 2.5-3.1 %) had it twice and 7.6 % (CI 7.1-8.1 %) once during the three rounds.

The risk of next-round PC diagnosis was 12.3-19.7 % following a FP result vs. 1.4-3.7 % following a screen-negative result, RR 3.6-9.9 (age-stratified risks, risk ratios and their CI presented in Table 2). There were 128 men who had FP in the first round and were diagnosed with PC in the second round – 78.1 % of the cancers were localized and low-grade (T1-2NxM0 and Gleason score < 7), 13.3 % were localized and Gleason score 7, and 8.6 % were advanced (T3-4NxM0 or TxNxM1 or Gleason score ≥ 8). In the men with a negative screening result in the first round and a screen-detected PC in the second round, corresponding numbers were 77.5 %, 13.2 % and 9.3 %. Similarly, there were 77 cancers in the third round in those men who had FP in the second round; 57.1 % were localized and low-grade (60.9 % in the first-round screen-negatives), 29.9 % were localized and Gleason score 7 (26.9 %), 13.0 % were advanced (12.2 %).

More than half of the men with one FP result had another in a subsequent round, while men with normal PSA levels had a 4.8-5.4 % risk of next-round FP result (Table 2). Men with a FP result were 1.5-2.0 times more likely to not participate in subsequent rounds compared to men with normal screening result (21.6-29.6 % vs. 14.0-16.7 %).

In the first round, moderately increased PSA concentration was associated with high probability of FP, whereas high PSA concentration (≥ 10 ng/ml) was associated with relatively high probability of PC (Table 3). Towards the third round, the probability of PC rose in the moderately increased PSA group and decreased in the high PSA group.

Information on use of medication for BPH was available for 23,319 men. The number of men who had used medication for BPH (finasteride or alpha-blockers or both) at first screen was 785 (3.8 % of participants, mean age 62.5 years vs. 60.1 years in men without BPH medication,), at second screen 1870 (10.1 % of participants, mean age 65.9 years vs. 63.8 years) and at third screen 460 (14.5 % of participants in the first year of the third round, mean age 67.1 years vs. 66.3 years). The men with BPH medication had roughly twice the risk for FP result compared to men without BPH medication: risk for FP was 14.0 % vs. 6.1 % in the first round, 13.4 % vs. 7.4 % in the second round and in the third round 11.7 % vs. 8.5 %. Age-adjusted first round RR was 1.9 (CI 1.5-2.2, $p < 0.001$), second round RR 1.6 (CI 1.4-1.8, $p < 0.001$) and third round RR 1.3 (CI 1.0-1.8, $p=0.04$). There was no increased risk for PC, as the risk in men with BPH medication vs. men without BPH medication was 2.4 % vs. 2.6 % in the first round (age-adjusted RR 0.8, CI 0.5-1.2, $p=0.20$), in the second round 3.8 % vs. 3.2 % (age-adjusted RR 1.0, CI 0.8-1.3, $p=0.99$) and in the third round 7.2 % vs. 4.7 % (age-adjusted RR 1.4, CI 1.0-2.0, $p=0.08$).

In the first round, men who were diagnosed with PC underwent on average 1.16 biopsies before diagnosis – that is, every sixth man underwent on average two biopsies. Men with a FP result had 1.30 biopsies in the follow-up time, i.e. every third man received two biopsies. In the second round, men with PC had 1.13 biopsies and men with FP 1.25 biopsies. These numbers decreased in the third round to 1.05 and 1.11, respectively. The maximum number of biopsies for a FP man was 7 (4 men) and for a man with PC 4 (1 man). Of the men with at least one FP result, 6.8 % had three or more biopsies.

Of the 1,331 men who had a FP result in the first round, 370 (27.8 %) developed a PC in the following 8-11 years (128 were diagnosed at the second screen, 28 at the third screen and 214 outside the screening protocol). Of these, 73.2 % were clinically localized and low-grade (T1-2NxM0 and Gleason < 7). Similarly, of the 1,489 FP men in the second round, 237 (15.9 %) were in the following 4-7 years diagnosed with PC (77 at the third screen, 160 outside the

screening protocol, 62.0 % localized and low-grade). Of 998 FP men in the third round 26 (2.6 %) developed PC later (in the following ≤ 3 years, 38.5 % localized and low-grade).

DISCUSSION

Our results show that false positive results affect every eighth man in repeated screening for prostate cancer with prostate-specific antigen even with a relatively high cut-off level of 4.0 ng/ml. More than a quarter of the men with FP results are subsequently diagnosed with PC, although most of these cancers are localized and low-grade and have similar characteristics as cancers in men with a previous negative screening test. More than half of these men have persistent high serum PSA levels resulting in repeated FP results and biopsies. They are also at high risk of dropping out of subsequent screening.

The Finnish Prostate Cancer Screening Trial is part of the ERSPC study. There are some differences between the ERSPC centres in e.g. mode of recruitment, screening interval, invitation procedures and the PSA threshold leading to biopsy. The Finnish trial is population-based and largest of the ERSPC centres. Population-based study design ensures good generalizability at the population level.

The ERSPC study recently showed preliminary mortality results indicating a 20 % relative decrease in mortality in the screening arm [Schröder et al, 2009]. This was the first evidence for benefits from screening for PC with PSA. However, as shown by the ERSPC trial, 1410 men would have to be offered screening and 48 PCs treated to prevent one prostate cancer death during a 9-year period. Also, the negative consequences of screening (adverse effects including overdiagnosis, overtreatment, costs) still need to be carefully evaluated to allow assessment of the balance between benefits and harms before evidence-based decision-making concerning provision of screening can be made. This analysis contributes to that requirement.

Our study presents a similar proportion of FP results per screening episode as a previous Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial study [Lafata et al, 2004], but provides also longer follow-up and information on the relation between FP results and several clinically important characteristics, such as PC, BPH medication, age and PSA level. Cumulative rates of FP results in repeated screening for several screening modalities were recently reported from the PLCO trial [Croswell et al, 2009]. The authors showed a risk of 10.4 % for at least one FP result in annual PSA screening during the 3-year screening period. We found a 12.5 % risk for at least one FP result in 2-3 successive screening rounds during 12 years follow-up, but the probability varied strongly with age. In the youngest age cohort (screened initially at age 55), the risk was 9.0 % and in the oldest men (first screen at age 67 with only two rounds of screening) 15.7 %. As previously noted, the incidence of PSA-elevating diseases other than PC (prostatitis, BPH) increases with age [Koskimäki et al, 1998; Rhodes et al, 1999; Wright et al, 2002] and explains the higher FP proportion in older men. This explanation is also consistent with the finding that men who used BPH medication had an increased risk for a FP result, despite the PSA-lowering effect of finasteride [Etzioni et al, 2005]. These men were also older than men without BPH medication.

In our study, a PSA threshold of 4.0 ng / ml was used. In addition, in 1996-1998 men with suspicious DRE finding and in 1999-2007 men with PSA 3.0-3.9 and free/total PSA ratio ≤ 16 % were referred. The PSA threshold was chosen in 1996 when the study began. A study from the Prostate Cancer Prevention Trial reported that 24.7 % of men in the placebo group with PSA 2.1-4.0 had PC when biopsied at the end of the study, albeit 50 % of these men were older than 71 years and all the cancers were stage T1 [Thompson et al, 2004]. In another study with younger subjects (50-65 years), 11.3 % (13/115) of men with PSA level 1.1-3.99 and F/T PSA ≥ 20 % had cancer in biopsy [Rowe et al, 2005]. In our study with men aged 55-71 years, the proportion of PC from those biopsied varied from 26.6 % to 32.9 %. If the PSA threshold had been lower, e.g. 2.1-3.9 ng/ml, the proportion of PC in the biopsied men would probably be smaller i.e. the downside of the expected improvement in sensitivity would be a decreased specificity.

A screen-detected cancer was defined as a PC detected within one year from the PSA test in a man with a screen-positive result. Based on this, we defined a FP as a screen-positive result with no PC diagnosis within one year from the PSA test (excluding men without biopsy). PSA predicts the development of PC by several years and there is no clear time as for the optimal definition of a FP result, but the proportion of de novo cases relative to those present at the screen can be anticipated to increase with time since PSA test. If we had extended our one year limit to e.g. three years, the number of FP results would have decreased by 86 (6.5 %), 88 (5.9 %) and 47 (4.7 %) men in the first, second and third rounds, respectively. These men were diagnosed with an interval cancer within 1-3 years from the PSA test. Since the proportion of these men out of all FP men was relatively small (4.7-6.5 %), using another definition would not be likely to materially affect our results.

In cancer screening, FP results are problematic for several reasons. Biopsies bring discomfort and often pain to the patient during the procedure [Mäkinen et al, 2002]. Waiting for the result is a psychological strain, which can have negative effects for at least a year even after a negative biopsy result [Fowler et al, 2006]. The economic impact of FP results has not been thoroughly analyzed, but these men seem to receive more follow-up interventions such as PSA testing and re-biopsies, which add the costs of screening [Lafata et al, 2004]. Biopsy – like any invasive procedure – involves risks for adverse health effects, such as bleeding, infection or abscess formation [Mäkinen et al, 2002], although these complications are not very common.

There is previous evidence that FP men undergo more follow-up testing and biopsies than men with normal PSA [Fowler et al, 2006]. Our results show that men with FP results receive more biopsies than the men who are diagnosed with PC. On average, every third FP man undergoes two biopsies within four years from the screen. It has been previously reported that the risk of clinically significant cancer decreases after second biopsy [Djavan et al, 2003]. Our study is likely to underestimate the average number of biopsies since we have no data on private sector visits and procedures and it is likely that some of the benign biopsies in the public sector are not reported to our database.

However, our findings indicate an increased risk for future PC with a history of one or several FP results. As many as 16 % of FP men were diagnosed with PC in the next round. Most of the PCs were not aggressive, but e.g. in the third round as many as 29.9 % of cancers were Gleason score 7 and 13 % were advanced (T3-4NxM0 or TxNxM1 or Gleason score 8). Of the first round FP men, almost a third was diagnosed with PC during the 8-11 follow-up years. The proportion of PC diagnoses among the men with FP results at the second and third

rounds were substantially lower (15.9 % and 2.6 %) - most likely because of a shorter follow-up. As previously mentioned, over 10 % of men over 50 years of age can be diagnosed with PC even with low PSA levels [Rowe et al, 2005]. Therefore, if men with a FP result receive more biopsies in the follow-up period than men with a negative screen, they could be more likely to receive a PC diagnosis due to more frequent biopsying. In addition, in 2002 we started using 10-12 core biopsies instead of sextant biopsies, which could increase chances of finding small, indolent lesions during the later follow-up period. Both these factors increase the PC risk in FP men.

When the men were stratified by serum PSA level, it was evident that at the first (prevalence) screen, high PSA level was clearly associated with PC and moderately increased serum PSA level with FP. At second and third (incidence) screens these differences evened out and the positive predictive value of high PSA for PC decreased. The most likely explanation for these trends is that at the first screen most of the high-PSA cancers were 'harvested' from the study population. Some of them were still detected at the second screen, but generally the cancers that produce high PSA were caught at prevalence screen and few such cases arose de novo between the screening rounds.

In the PLCO trial, men with a FP result were almost twice more likely to decline subsequent screening compared to men with a negative screening result [Ford et al, 2005]. Our results are similar, with risk ratios varying from 1.5 to 2.0. There might be several reasons behind this. The FP men could decide not to participate because of the unpleasant experience of unnecessary biopsy procedure and the anxiety related to the fear of PC diagnosis. On the other hand, a FP man could sense relief after a benign biopsy and deem unnecessary to participate in the next screening round. Also, receiving a positive screening result without a confirmed PC diagnosis may erode man's perception of effectiveness of screening.

These findings emphasize the paradoxical problem of the FP results in PC screening. On one hand, FP men frequently have persistent high PSA levels (> 50 % chance of having another FP result in the next round) and undergo several biopsies. On the other hand, they are more likely to be diagnosed with PC, either due to biological processes or more active diagnostic procedures. New approaches are urgently needed for improved risk stratification among these men i.e. to predict which of them may harbour a clinically significant PC, which may have an insignificant indolent PC and which may have other factors underlying the elevated PSA level.

There is one weakness in our study. In some cases, the follow-up time after the third screen was relatively short (≤ 3 years), as the last men were screened in the end of 2007 and follow-up ended in 2007. Therefore some post-screening cancers were lacking for the last screening cohort. However, we believe that the strengths of this study well outweigh this weakness.

In conclusion, we present data from a prospective randomized controlled prostate cancer screening study spanning 12 years and three screening rounds. We have analyzed the FP screening results during these rounds and calculated that every eighth man screened is subject to a FP result at least once in repeat screening. The men who receive FP results are likely to have a subsequent FP result(s) later if screened again. Also, these men commonly drop out of subsequent screening rounds. This poses a difficult equation, as men with FP results are in an increased risk of being diagnosed with a prostate cancer later. More research is needed to balance the sensitivity and specificity of prostate cancer screening to minimize the proportion of false positive results.

ACKNOWLEDGEMENTS

The Finnish randomized trial has been supported by the Academy of Finland (grant no 123054), The Competitive Research Funding of the Pirkanmaa Hospital District and Finnish Cancer Organizations.

REFERENCES

Andriole GL, Crawford ED, Grubb RL 3rd, Buys SS, Chia D, Church TR, Fouad MN, Gelmann EP, Kvale PA, Reding DJ, Weissfeld JL, Yokochi LA, O'Brien B, Clapp JD, Rathmell JM, Riley TL, Hayes RB, Kramer BS, Izmirlian G, Miller AB, Pinsky PF, Prorok PC, Gohagan JK, Berg CD; PLCO Project Team (2009) Mortality results from a randomized prostate-cancer screening trial. N Engl J Med 360:1310-9.

Catalona WJ, Richie JP, Ahmann FR, Hudson MA, Scardino PT, Flanigan RC, deKernion JB, Ratliff TL, Kavoussi LR, Dalkin BL (1994) Comparison of digital rectal examination and serum prostate-specific antigen in the early detection of prostate cancer: Results of a multicenter clinical trial of 6,630 men. J Urol 151:1283-90.

Croswell JM, Kramer BS, Kreimer AR, Prorok PC, Xu JL, Baker SG, Fagerstrom R, Riley TL, Clapp JD, Berg CD, Gohagan JK, Andriole GL, Chia D, Church TR, Crawford ED, Fouad MN, Gelmann EP, Lamerato L, Reding DJ, Schoen RE (2009) Cumulative incidence of false-positive results in repeated, multimodal cancer screening. Ann Fam Med 7:212-222.

Djavan B, Remzi M, Marberger M (2003) When to biopsy and when to stop biopsying. Urol Clin North Am 30:253–62.

Etzioni RD, Howlader N, Shaw PA, Ankerst DP, Penson DF, Goodman PJ, Thompson IM (2005) Long-term effects of finasteride on prostate specific antigen levels: results from the prostate cancer prevention trial. J Urol 174:877-81.

Ford ME, Havstad SL, Demers R, Johnson CC (2005) Effects of false-positive prostate cancer screening results on subsequent prostate cancer screening behavior. Cancer Epidemiol Biomarkers Prev 14:190-194.

Fowler FJ Jr, Barry MJ, Walker-Corkery B, Caubet JF, Bates DW, Lee JM, Hauser A, McNaughton-Collins M (2006) The Impact of a suspicious prostate biopsy on patients' psychological, socio-behavioral and medical care outcomes. J Gen Intern Med 21:715-721.

Koskimäki J, Hakama M, Huhtala H, Tammela TL (1998) Prevalence of lower urinary tract symptoms in Finnish men: a population-based study. Br J Urol 81:364-9.

Lafata J, Simpkins J, Lamerato L, Poisson L, Divine G, Johnson C (2004) The economic impact of false-positive cancer screens. Cancer Epidemiol Biomarkers Prev 13:2126–32.

Merrill RM, Feuer EJ, Warren JL, Schussler N, Stephenson RA (1999) Role of transurethral resection of the prostate in population-based prostate cancer incidence rates. Am J Epidemiol 150:848-860.Määttänen L, Hakama M, Tammela TLJ, Ruutu M, Ala-Opas M, Juusela H, Martikainen P, Stenman UH, Auvinen A (2007) Specificity of serum prostate-specific antigen determination in the Finnish prostate cancer screening trial. Br J Cancer 96:56-60.

Mäkinen T, Auvinen A, Hakama M, Stenman UH, Tammela TLJ (2002) Acceptability and compliations of prostate biopsy in population-based PSA screening versus routine clinical practice: a prospective, controlled study. Urology 60:846-850.

Parkin DM, Bray F, Ferlay J, Pisani P (2005) Global Cancer statistics, 2002. CA Cancer J Clin 55:74-108.

Rhodes T, Girman CJ, Jacobsen SJ, Roberts RO, Guess HA, Lieber MM (1999) Longitudinal prostate growth rates during 5 years in randomly selected community men 40 to 79 years old. J Urol 161:1174-9.

Rowe EW, Laniado ME, Walker MM, Patel A (2005) Prostate cancer detection in men with a 'normal' total prostate-specific antigen (PSA) level using percentage free PSA: a prospective screening study. BJU Int 95:1249-52.

Schröder FH, van der Maas P, Beemsterboer P, Kruger AB, Hoedemaeker R, Rietbergen J, Kranse R (1998) Evaluation of the digital rectal examination as a screening test for prostate cancer. Rotterdam section of the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst 90:1817-23.

Schröder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, Kwiatkowski M, Lujan M, Lilja H, Zappa M, Denis LJ, Recker F, Berenguer A, Määttänen L, Bangma CH, Aus G, Villers A, Rebillard X, van der Kwast T, Blijenberg BG, Moss SM, de Koning HJ, Auvinen A; ERSPC Investigators (2009) Screening and prostate cancer mortality in a randomized European study. N Engl J Med 360:1320-8.

Stenman UH, Finne P, Zhang W, Leinonen J (2000) Prostate-specific antigen and other prostate cancer markers. Urology 56:893-8.

Teppo L, Pukkala E, Lehtonen M (1994) Data quality and quality control of a population-based cancer registry. Experience in Finland. Acta Oncol 33:365-9.

Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, Minasian LM, Ford LG, Lippman SM, Crawford ED, Crowley JJ, Coltman CA Jr (2004) Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. N Engl J Med 350:2239-46.

Welch HG, Albertsen PC (2009) Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986-2005. J Natl Cancer Inst 101:1325-9.Wright EJ, Fang J, Metter EJ, Partin AW, Landis P, Chan DW, Carter HB (2002) Prostate specific antigen predicts the long-term risk of prostate enlargement: results from the Baltimore Longitudinal Study of Aging. J Urol 167:2484-7.

Table 1. Screening results by age and screening round.

	Screen negative, N (%)	False positive, N (%)	Prostate cancer, N (%)	Not biopsied, N (%)	Total, N
Round I					
55 yrs	6,146 (95.2)	210 (3.3)	76 (1.2)	21 (0.3)	6,453
59 yrs	5,136 (92.5)	274 (4.9)	130 (2.3)	12 (0.2)	5,552
63 yrs	4,143 (87.6)	397 (8.4)	153 (3.2)	35 (0.7)	4,728
67 yrs	3,387 (83.5)	450 (11.1)	184 (4.5)	35 (0.9)	4,056
Total	18,812 (90.5)	1,331 (6.4)	543 (2.6)	103 (0.5)	20,789
Round II					
59 yrs	5,700 (92.4)	313 (5.1)	115 (1.9)	39 (0.6)	6,167
63 yrs	4,464 (88.0)	401 (7.9)	158 (3.1)	48 (0.9)	5,071
67 yrs	3,426 (84.9)	372 (9.2)	181 (4.5)	58 (1.4)	4,037
71 yrs	2,719 (81.5)	403 (12.1)	159 (4.8)	57 (1.7)	3,338
Total	16,309 (87.6)	1,489 (8.0)	613 (3.3)	202 (1.1)	18,613
Round III					
63 yrs	4,833 (89.5)	352 (6.5)	157 (2.9)	59 (1.1)	5,401
67 yrs	3,618 (86.5)	334 (8.0)	164 (3.9)	66 (1.6)	4,182
71 yrs	2,645 (83.8)	312 (9.9)	134 (4.2)	66 (2.1)	3,157
Total	11,096 (87.1)	998 (7.8)	455 (3.6)	191 (1.5)	12,740
Total	46,217 (88.6)	3,818 (7.3)	1,611 (3.1)	496 (1.0)	52,142

Table 2. Risks for next-round prostate cancer, FP result and non-participation after a previous round FP result and negative screening result.

	Risk for PC after FP result, %	Risk for PC after negative screen, %	RR (95 % CI)	Risk for FP result after FP result, %	Risk for FP after negative screen, %	RR (95 % CI)	Risk for non-participation after FP result, %	Risk for non-participation after negative screen, %	RR (95 % CI)4
Round 1 / Round 2									
55 yrs	14.1	1.4	9.9 (6.1-16.1)	50.0	4.1	12.2 (9.8-15.1)	27.8	16.4	1.7 (1.3-2.2)
59 yrs	15.2	2.4	6.3 (4.3-9.4)	64.0	5.5	11.6 (9.8-13.7)	21.6	14.3	1.5 (1.2-1.9)
63 yrs	19.7	3.2	6.2 (4.5-8.5)	53.6	6.0	9.0 (7.5-10.7)	26.8	14.5	1.8 (1.5-2.2)
67 yrs	14.2	3.7	3.9 (2.7-5.5)	62.5	7.1	8.8 (7.4-10.4)	27.3	14.0	1.9 (1.6-2.3)
Total	16.0	2.5	6.5 (5.4-7.8)	58.3	5.4	10.7 (9.8-11.7)	26.0	15.0	1.7 (1.6-1.9)
Round 2 / Round 3									
59 yrs	12.9	2.1	6.2 (4.0-9.6)	52.0	4.3	12.0 (9.8-14.6)	29.3	16.7	1.8 (1.4-2.1)
63 yrs	12.3	3.4	3.6 (2.4-5.3)	59.1	4.5	13.3 (11.0-16.0)	29.6	14.6	2.0 (1.7-2.4)
67 yrs	13.6	3.0	4.5 (3.0-6.7)	58.7	5.9	10.0 (8.3-12.1)	29.3	16.3	1.8 (1.5-2.2)
Total	12.9	2.8	4.7 (3.7-5.9)	57.0	4.8	12.0 (10.7-13.4)	29.4	15.9	1.8 (1.7-2.1)

Table 3. Proportions of men with false positive results and prostate cancer by serum PSA concentration in the three rounds of the trial.

	PSA, ng/ml	Men, N (%)	Men biopsied, N (%)	False positives, N (%)	Prostate cancer, N (%)
Round 1	≤ 2.9 or 3.0-3.9 and aux. test -	18,812 (90.5)	-	-	-
	3.0-3.9 and aux. test +	149 (0.7)	142 (95.3)	105 (73.9)	37 (26.1)
	4.0-9.9	1,527 (7.3)	1,440 (94.3)	1,110 (77.1)	330 (22.9)
	≥ 10.0	301 (1.4)	292 (97.0)	116 (39.7)	176 (60.3)
	Total	20,789	1,874 (9.0)	1,331 (6.4)	543 (2.6)
Round 2	≤ 2.9 or 3.0-3.9 and aux. test -	16,309 (87.6)	-	-	-
	3.0-3.9 and aux. test +	232 (1.2)	215 (92.7)	147 (68.4)	68 (31.6)
	4.0-9.9	1,819 (9.8)	1,653 (90.9)	1,203 (72.8)	450 (27.2)
	≥ 10.0	253 (1.4)	234 (92.5)	139 (59.4)	95 (40.6)
	Total	18,613	2,102 (11.3)	1,489 (8.0)	613 (3.3)
Round 3	≤ 2.9 or 3.0-3.9 and aux. test -	11,096 (87.1)	-	-	-
	3.0-3.9 and aux. test +	160 (1.3)	142 (88.8)	88 (62.0)	54 (38.0)
	4.0-9.9	1,328 (10.4)	1,167 (87.9)	813 (69.7)	354 (30.3)
	≥ 10.0	156 (1.2)	144 (92.3)	97 (67.4)	47 (32.6)
	Total	12,740	1,453 (11.4)	998 (7.8)	455 (3.6)
	Auxiliary test: 1996-1998 digital rectal examination, 1999-2007 free/total PSA ratio (cut-off 16 %)				

II

Kilpeläinen TP, Tammela TL, Roobol M, Hugosson J, Ciatto S, Nelen V, Moss S, Määttänen L, Auvinen A. False-positive screening results in the European randomized study of screening for prostate cancer. Eur J Cancer. 2011;47:2698-2705.

EUROPEAN JOURNAL OF CANCER 47 (2011) 2698–2705

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

False-positive screening results in the European randomized study of screening for prostate cancer

Tuomas P. Kilpeläinen [a,b,*], *Teuvo L.J. Tammela* [a], *Monique Roobol* [c], *Jonas Hugosson* [d], *Stefano Ciatto* [e], *Vera Nelen* [f], *Sue Moss* [g], *Liisa Määttänen* [h], *Anssi Auvinen* [b]

[a] *Dept. of Urology, Tampere University Hospital and University of Tampere, Tampere, Finland*
[b] *Tampere School of Public Health, University of Tampere, Tampere, Finland*
[c] *Dept. of Urology, Erasmus, MC, Rotterdam, The Netherlands*
[d] *Dept. of Urology, Sahlgrenska University Hospital, Göteborg, Sweden*
[e] *Dept. of Diagnostic Medical Imaging, ISPO, Firenze, Italy*
[f] *Provinciaal Instituut voor Hygiëne, Antwerp, Belgium*
[g] *Cancer Screening Evaluation Unit, Institute of Cancer Research, Surrey, UK*
[h] *Finnish Cancer Registry, Helsinki, Finland*

ARTICLE INFO

Article history:
Available online 23 July 2011

Keywords:
Mass screening
Prostatic neoplasms
PSA
Randomized controlled trials
Sensitivity and specificity

ABSTRACT

Background: Screening for prostate cancer (PC) with prostate-specific antigen (PSA) has been shown to decrease mortality, but has adverse effects, such as false-positive (FP) screening results. We describe the frequency of FP results and assess their relation to subsequent screening attendance, test results and prostate cancer risk in a large randomized trial.

Materials and methods: We included data from five centres of the European Randomized Study of Screening for Prostate Cancer, altogether over 61,000 screened men. Men were screened with PSA test at a 2–7 year interval depending on the centre; PSA cut-off was 3.0–4.0 ng/ml. A positive screen with no histologically confirmed PC in biopsy within 1 year was defined as an FP result.

Results: Of the 61,604 men who were screened at least once, 17.8% had one or more FP result(s). Almost 20% of men who participated at all screening rounds had one or more FP result(s). More than half of the men with an FP result had another FP if screened again. Men with FP results had a fourfold risk of PC at subsequent screen (depending on the round, 10.0% versus 2.6–2.7% of men with negative screen, risk ratio 3.8–3.9). The PCs following an FP result were in 92.8% of cases localised and low-grade versus 90.4% following a screen-negative result.

Conclusions: Our results show that FP results are common adverse effects in PC screening, as they affect at least one in six screened men. False-positive men are more prone to be diagnosed with PC but are also likely to have consistently high PSA levels.

1. Introduction

The benefits of screening for prostate cancer (PC) with prostate-specific antigen (PSA) test are becoming clearer with evidence from the European Randomized Study of Screening for Prostate Cancer (ERSPC), which has shown a 20% relative decrease in mortality in the screening arm.[1] However, the adverse effects and cost-effectiveness of screening need to be

* *Corresponding author:* Address: Tampere School of Public Health and University of Tampere, FIN-33014 Tampere, Finland. Fax: +358 335516057.
E-mail address: tuomas.kilpelainen@uta.fi (T.P. Kilpeläinen).
0959-8049/$ - see front matter
doi:10.1016/j.ejca.2011.06.055

thoroughly investigated before decisions regarding population-based screening can be made.

Serum PSA is an organ-specific marker that may be affected by any prostatic disease. Therefore, as all screening tests, it is not perfect in sensitivity or specificity. Undetected disease constitutes a false negative (FN) finding and a positive screening result in the absence of disease is a false positive (FP) result (Table 1). The challenge in PC screening is to define and predict the disease status based on the PSA test, as not all subjects can undergo the diagnostic test, the prostatic biopsy. Even the biopsy has uncertainties: first, the needle biopsy provides only a small sample of the prostate tissue potentially missing the cancer lesion and second, the presence of malignant tissue does not necessarily mean clinically significant PC (resulting in overdiagnosis). Overdiagnosis and overtreatment occur when PCs that would not have been diagnosed in the absence of screening are detected by screening and treated.[2,3]

In addition, screening for PC with PSA has relatively high FP rate, or conversely, low specificity.[4] Previously, the results from the Finnish component of the ERSPC trial have shown that 12.5% of the screened men (at 4-year interval) had an FP result at least once during three screening rounds.[5] Similarly, 10.4% of men in the Prostate, Lung, Colorectal and Ovarian cancer screening trial (PLCO) had an FP result during four PSA tests and 3 years of follow-up.[6]

We present the proportion of FP results during three screening rounds in five centres of the ERSPC trial: Belgium, Finland, Italy, the Netherlands and Sweden – with more than 61,000 screened men. We also investigated subsequent screening compliance, PC risk and repeated FP result(s).

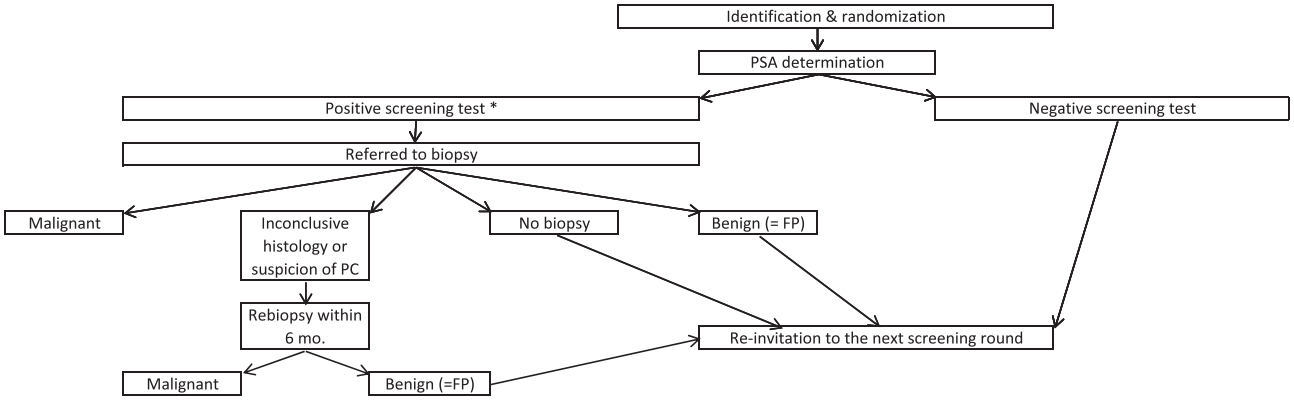
2. Materials and methods

The ERSPC trial is a multicentre study in eight European countries. In this study, we analysed data from five centres: Belgium, Finland, Italy, the Netherlands and Sweden. These five centres had data from at least three screening rounds and the largest numbers of men.

There was some variation between the centres in the screening protocol (Graph 1), mainly in the screening interval, PSA threshold, age of screened men and the mode of recruitment (Table 2). The screening protocols in the ERSPC centres have been described in detail elsewhere.[7–10] Men with a PC diagnosis and those who had emigrated from the study region were no longer invited. Men who chose not to participate were re-invited to the following rounds, except in the Netherlands. Due to lack of funding, the first screening interval in Belgium was delayed up to seven years. The Swedish centre used biennial screening and therefore had six screening rounds. In Sweden, the men with PSA <1.0 ng/ml in round 2

Table 1 – Relationship between terms regarding sensitivity and specificity.

		DISEASE		
		Positive	Negative	
TEST	Positive	True positive (TP)	False positive (FP)	Positive predictive value = TP / (TP + FP)
	Negative	False negative (FN)	True negative (TN)	Negative predictive value = TN / (FN+TN)
		Sensitivity TP / (TP+FN)	Specificity TN / (FP + TN)	



* Defined by criteria in each center, see text for details.
PC = prostate cancer; FP = false positive

Graph 1 – A general flow chart of the screening process.

Table 2 – Characteristics of the screening protocols in the ERSPC centres (screening interval, PSA cut-off, age range, recruitment mode and time of data collection).

	Interval, years	PSA cut-off, ng/ml	Age at entry (mean age)	Recruitment	Screening began	Followed up to
Belgium[a]	4–7	3.0	55–75 (64.3)	Volunteer	Jun. 1991–Dec. 2003	31.12.2007
Finland[b]	4	4.0	55–67 (60.1)	Population	Jan. 1996–Jan. 1999	31.5.2008
Italy[c]	4	4.0	55–71 (62.4)	Population	Oct. 1996–Oct. 2000	31.7.2008
Netherlands[d]	4	3.0/4.0	55–75 (63.6)	Volunteer	Nov. 1993–Mar. 2000	31.8.2008
Sweden[e]	2	3.0	50–64 (56.3)	Population	Dec. 1994	30.6.2008

PSA, prostate-specific antigen; DRE, digital rectal examination; TRUS, transrectal ultrasound.

[a] In the first round, PSA, DRE and TRUS to all. In the second round, PSA and DRE to all. PSA cut-off for biopsy 10 ng/ml in 1992–1994, 4 ng/ml in 1995–1998 and 3 ng/ml from 1999 onwards.

[b] For PSA 3.0–3.9 ng/ml: DRE in 1996–1998; DRE replaced by free/total PSA ratio with cut-off 16% from 1999 onwards.

[c] For PSA 2.5–4.0 ng/ml: DRE and TRUS (biopsy if suspicious). Biopsy for all with PSA >4.0 ng/ml.

[d] DRE and TRUS initially to all men 1993–1995; those with PSA >1.0, 1995–1997, abandoned 1997 onwards. PSA cut-off 4.0 ng/ml was lowered to 3.0 ng/ml in May 1997.

[e] No ancillary test.

were not invited to round 3, but were subsequently invited to rounds 4–6.

Incident PC cases were identified from the trial database and local/national cancer registries. Only screen-detected PCs were analysed in this study (i.e. no interval cancers). An FP result was defined as a positive screening result without a cancer diagnosis in subsequent histological examination within 1 year from the screen. Men who did not undergo biopsy were not regarded FP. Men with PC diagnosis after 1 year in e.g. a re-biopsy were classified as with interval cancers. T3-4N0M0 or T1-4N0-2M1 or Gleason score $\geqslant 8$ cancer was defined as aggressive cases.

The study protocols were reviewed and approved by appropriate ethical committees in each participating country. Written informed consent was obtained from the screened men. In Belgium and the Netherlands, men were randomised after the informed consent was obtained due to legislative reasons. In Finland, Italy and Sweden, men were randomised to screening arm and control arm prior to informed consent and only men in the screening arm were contacted.

The 95% confidence intervals (CIs) for risks and proportions were calculated on the basis of basic standard error formulae. A generalised linear model for binomial distribution with a logarithmic link function was used to calculate age-adjusted risk ratios (RRs) and their 95% CIs. Age-standardisation for the prevalences of results was done using the entire study population as reference group, based on mean age group proportions. Spearman's correlation coefficient for proportion of PC and FP results was calculated summing up all rounds and centres. Statistical analyses were performed using Stata 8.2 (StataCorp, College Station, TX, USA).

3. Results

Overall, 61,604 men screened in the five centres of ERSPC trial were analysed in this study. Of them, 22,068 (35.8%) men participated in all rounds (three rounds, except in Sweden six rounds). Altogether 4733 PCs were detected by screening, yielding a detection probability of 3.4% (95% CI 3.2–3.5) in the first round, 3.4% (3.2–3.5) in the second round and 3.6% (3.4–3.8) in the third round (Table 3).

The proportion of FP results was 10.2% (95% CI 10.0–10.5; specificity 89.8%) in the first round, 11.0% (10.7–11.3; specificity 89.0%) in the second round and 11.1% (10.7–11.5; specificity 88.9%) in the third round. There was substantial variation between the centres (Table 3). The proportion of FP results per round varied from 3.5% to 20.6% depending on age, with a higher proportion in older ages (Table 4). The proportion of PC also increased with age. Spearman's correlation coefficient was 0.78 ($p < 0.001$) between the proportion of PC and FP results per round.

Altogether 10,972 men (17.8% of those who were screened at least once) had one or more FP result (Table 5). The proportion of men with FP result(s) varied by the centre from 11.1% in Italy to 26.4% in the Netherlands. The majority (74.7%) of the men with FP result(s) had only one FP result. Of the 22,068 men who participated in all (three to six) rounds, 19.0% had one or more FP result(s).

An FP result increased the risk for a next-round FP, with approximately 50% of the men having another FP result if they participated in the following round (Table 6). A negative screen carried a risk of 6.2–7.7% for an FP result in the next round (RR 6.5–8.6 by round); the RR relative to initially screen-negative men varied by the centre 2.5–15.0. A quarter of the men with FP results dropped out of the subsequent screening round, which was 1.6-fold following the first round and 1.5-fold after the second screen compared to the screen-negative men. RRs varied by the centre from 0.7 (Sweden) to 2.1 (Finland) (Table 6).

The absolute risk of a screen-detected PC in the next round following a previous FP (at first or second round) was 10%, which was roughly fourfold (ranging from 1.7 to 14.4 depending on the centre and screening round) compared to men with a negative screen in the previous round (Table 6). The positive predictive value (PPV) was highest (22.0% in the first, 22.7% in the second and 25.0% in the third round) among those screen-positive men who had not undergone a biopsy before. The PPV for the previous-round of FP men was 14.3% and 13.6% in the second and third rounds, respectively.

A total of 681 PCs were detected at screening following an FP result. Of those, 92.8% ($N = 632$) were non-aggressive and

Table 3 – The number and proportion of results in the five centres of the trial during 3–6 rounds.

	Participation proportion, N (%)	Screen-negative, N (%)	Screen-positives, N (%)		
			False-positive	Not biopsied	Prostate cancer
All centres					
Round 1	56,064/72,210 (77.6)	47,461 (84.7)	5722 (10.2)	985 (1.8)	1896 (3.4)
Round 2	42,884/61,003 (70.3)	35,711 (83.3)	4732 (11.0)	992 (2.3)	1449 (3.4)
Round 3	27,835/42,248 (65.9)	22,929 (82.4)	3090 (11.1)	819 (2.9)	997 (3.6)
Belgium					
Round 1	4562/5178 (87.0*)	3916 (88.1*)	325 (6.1*)	214 (3.6*)	107 (2.2*)
Round 2	1987/3430 (56.8*)	1550 (80.9*)	237 (9.4*)	99 (5.8*)	101 (3.8*)
Round 3	718/1336 (51.3*)	593 (85.0*)	62 (8.5*)	49 (5.2*)	14 (1.4*)
Finland					
Round 1	20,789/30,197 (69.4*)	18,812 (90.0*)	1332 (7.0*)	102 (0.6*)	543 (2.8*)
Round 2	18,613/26,324 (71.2*)	16,309 (86.0*)	1467 (8.8*)	224 (1.4*)	613 (3.8*)
Round 3	12,739/18,376 (69.4*)	11,095 (86.3*)	978 (8.2*)	198 (1.7*)	468 (3.9*)
Italy					
Round 1	4908/5696 (85.4*)	4300 (88.0*)	377 (7.1*)	142 (3.3*)	89 (1.7*)
Round 2	4499/5607 (80.7*)	3942 (87.9*)	267 (5.9*)	217 (4.7*)	73 (1.5*)
Round 3	3292/5533 (62.2*)	2844 (86.1*)	145 (4.4*)	269 (8.5*)	34 (1.0*)
Netherlands					
Round 1	19,950/21,175 (94.3*)	15,240 (79.2*)	3225 (14.6*)	470 (1.9*)	1015 (4.3*)
Round 2	12,525/16,163 (77.4*)	9259 (74.1*)	2360 (18.7*)	355 (2.8*)	551 (4.4*)
Round 3	7711/9799 (72.5*)	5848 (75.8*)	1326 (17.2*)	217 (2.8*)	320 (4.1*)
Sweden					
Round 1	5855/9964 (57.0*)	5193 (85.1*)	463 (10.6*)	57 (1.1*)	142 (3.3*)
Round 2	5260/9479 (58.7*)	4651 (87.7*)	401 (8.0*)	97 (2.0*)	111 (2.3*)
Round 3	3375/7204 (63.0*)	2549 (76.8*)	579 (16.5*)	86 (2.4*)	161 (4.4*)
Round 4	4622/7851 (58.9)	3888 (84.1)	496 (10.7)	105 (2.3)	133 (2.9)
Round 5	4114/6674 (61.6)	3499 (85.1)	435 (10.6)	69 (1.7)	111 (2.7)
Round 6	3475/5688 (61.1)	2773 (79.8)	467 (13.4)	88 (2.5)	147 (4.2)

Proportions marked with an asterisk (*) are age-standardised to the mean age distribution of the round in all the centres.

Table 4 – Age-stratified aggregate results from the three rounds of the trial.

	Age at screen, years				
	<55	55–59	60–64	65–69	⩾70
Round 1					
Screen-negative	4579 (94.9)	18,299 (90.7)	12,614 (83.9)	9163 (77.7)	2806 (66.4)
False-positive	171 (3.5)	1299 (6.4)	1674 (11.1)	1706 (14.5)	872 (20.6)
Not biopsied	21 (0.4)	167 (0.8)	260 (1.7)	293 (2.5)	244 (5.8)
Screen-detected PC	55 (1.1)	414 (2.1)	491 (3.3)	634 (5.4)	302 (7.1)
Total	4826	20,179	15,039	11,796	4224
Round 2					
Screen-negative	792 (94.9)	8402 (91.4)	11,746 (84.9)	8741 (79.4)	6030 (75.3)
False-positive	34 (4.1)	552 (6.0)	1402 (10.1)	1470 (13.4)	1274 (15.9)
Not biopsied	7 (0.8)	82 (0.9)	278 (2.0)	326 (3.0)	299 (3.7)
Screen-detected PC	2 (0.2)	160 (1.7)	410 (3.0)	473 (4.3)	404 (5.0)
Total	835	9196	13,836	11,010	8007
Round 3					
Screen-negative	38 (84.4)	910 (82.2)	7450 (86.1)	8612 (81.9)	5919 (78.7)
False-positive	6 (13.3)	141 (12.7)	766 (8.9)	1162 (11.1)	1015 (13.5)
Not biopsied	0 (0.0)	16 (1.4)	166 (1.9)	334 (3.2)	303 (4.0)
Screen-detected PC	1 (2.2)	40 (3.6)	269 (3.1)	401 (3.8)	286 (3.8)
Total	45	1107	8651	10,509	7523

6.3% ($N = 43$) aggressive (missing information for 0.9% or six cases). Of the 1725 PCs following a screen-negative result, 90.4% ($N = 1560$) were non-aggressive and 7.8% ($N = 134$) aggressive (1.8%, $N = 31$ with missing information). The difference in proportion of aggressive cancers among those following an FP was statistically non-significantly lower compared with screen-detected cases subsequent to a screen-negative result (6.3% versus 7.8%, $p = 0.11$).

Table 5 – The prevalence of false-positive (FP) results in five centres of the trial.

	Men participating at least once, *N*	Men with FP(s), % (*N*)[a]	1 FP, % (*N*)	2 FPs, % (*N*)	3 FPs, % (*N*)	Men participating every round, *N*	Men with FP(s), *N* (%)[a]
All centres	61,604	17.8 (10,972)	74.7 (7752)	20.1 (2089)	5.2 (538)	22,068	19.0 (4186)
Belgium	4677	11.0 (569)	90.7 (516)	9.0 (51)	0.4 (2)	584	15.6 (105)
Finland	23,771	13.0 (2934)	75.2 (2207)	20.8 (611)	4.0 (116)	10,326	11.9 (1184)
Italy	5696	10.5 (635)	78.7 (500)	18.3 (116)	3.0 (19)	2597	9.0 (286)
Netherlands	19,950	26.1 (5266)	74.3 (3912)	20.2 (1063)	5.5 (291)	7711	27.8 (2228)
Sweden[b]	7510	22.3 (1568)	55.9 (876)	22.1 (347)	12.4 (195)	850 2112[c]	44.9 (383) 20.8 (442)[c]

Sweden[b]	4 FPs %, (N)	5 FPs (%)	6 FPs (%)
	5.2 (81)	3.3 (52)	1.1 (17)

[a] Age-standardised proportion.
[b] Sweden has six screening rounds.
[c] Excluding the third round of the Swedish trial (the men with PSA <1.0 ng/ml in round 2 were not invited to round 3, but were subsequently invited to rounds 4–6).

If the PSA threshold would have been 4.0 ng/ml in all the centres, the proportion of FP results would have decreased from 17.8% to 11.7% (10,972 versus 7182). However, fewer PCs would have been detected: 3481 instead of 4733 (91.5% of these PCs would have been non-aggressive, 6.7% aggressive, 1.8% unknown). In Belgium, the proportion of FP results would have been 6.4% (instead of 11.0%); in Finland 12.0% (13.0%); in Italy 9.3% (10.5%); in the Netherlands 12.4% (26.1%) and in Sweden 14.0% (22.3%).

4. Discussion

The results from five centres of the ERSPC trial show that false-positive screening results affect one in six screening participants during the course of the screening programme. Almost 20% of the men who participate in every (three to six) screening round encounter an FP result at least once. Men with FP results are often diagnosed with PC in the next round and more than half have another FP result if re-screened. The men with FP results are also more likely to drop out of the subsequent screening rounds. Our results also show that there are marked differences in the prevalence of FP results between the ERSPC centres, most likely due to differences in PSA threshold, but this could also reflect the underlying PC risk.

The ERSPC trial was launched in the early 1990s to assess whether screening for PC with PSA decreases PC mortality. The early results have shown a relative mortality decrease of 20% in the screening versus control arm.[1] When adjusted for non-attendance and contamination, the relative decrease was approximately 31%.[11] However, despite these promising results, the adverse effects and cost-effectiveness of screening need to be evaluated thoroughly if screening is to be recommended in the future.

False-positive screening results represent one aspect of the adverse effects of PC screening, in addition to overdiagnosis and overtreatment. FP results can be problematic for several reasons, even though the prostate biopsy as such seldom results in complications.[12,13] Waiting for the biopsy and afterwards the result of the biopsy can be psychologically straining to the patient, even if the biopsy eventually turns out to be negative.[14] Men with FP results commonly undergo repeated follow-up biopsies, which increase the costs of screening and could reduce the compliance.[5,14,15]

We defined an FP result as a screen-positive result without a PC diagnosis in biopsy within a year from the PSA test. The one-year time limit was adopted to ensure comparability between the centres. If this time limit is extended, some missed PCs are detected in FP men (rendering them true positive) but also PCs arising de novo after screening become more common. The definition of an FP result is problematic, however, as elevated PSA resulting in FP may indicate a PC missed in biopsy (i.e. a true positive, which overestimates the FP prevalence) or, a biopsy may result in the diagnosis of an indolent PC (and the screening test could be interpreted as FP in the sense that no clinically significant disease was diagnosed, with underestimation of the FP frequency).

High prevalence of FP results is a well-known issue with screening for PC with PSA and active search for a better screening tool has been ongoing to increase specificity. There is evidence on the usefulness of the free/total PSA ratio,[16] especially when combined with PSA density and digital rectal examination in multivariate regression models (reduction in FP results was 22%).[17] PSA velocity has not been proven very effective in increasing specificity.[18,19] New biomarkers (e.g. kallikrein-related peptidase 2, proPSA, nicked PSA, p2PSA) have shown potential, but are not yet in widespread clinical use.[20–22]

Our aggregate results from the ERSPC trial confirm that FP results are common in PC screening – one in six men have at least one FP result during the screening protocol. Three quarters of them have only one FP, but 25% have two or three. There is, however, much variation between the centres in the risk of FP result(s).

Why is the prevalence of FP results so different between the centres? One explanation could be age, as the frequency of FP results increases with age. This is most likely due to other PSA-elevating prostatic diseases than PC, such as benign prostatic hyperplasia (BPH) and prostatitis, which become more common with age.[23–25] Most of our results are,

Table 6 – Risks for subsequent round prostate cancer (PC), false-positive result (FP) and non-participation after previous round FP result versus negative screening result.

	Risk for PC after FP result (%)	Risk for PC after negative screen (%)	RR (95% CI)	Risk for FP after FP result (%)	Risk for FP after negative screen (%)	RR (95% CI)	Risk for non-participation after FP result (%)	Risk for non-participation after negative screen (%)	RR (95% CI)
All centres									
Round 1/2	10.0	2.7	3.7 (3.3–4.2)	50.1	7.7	6.5 (6.2–6.8)	26.8	17.2	1.6 (1.5–1.6)
Round 2/3	10.0	2.6	3.9 (3.3–4.5)	53.0	6.2	8.5 (8.0–9.0)	27.8	18.6	1.5 (1.4–1.6)
Belgium									
Round 1/2	7.9	4.6	1.7 (0.9–3.4)	26.7	10.6	2.5 (1.8–3.6)	49.0	36.5	1.3 (1.2–1.6)
Round 2/3	7.8	1.3	6.0 (2.0–18.3)	39.1	4.8	8.1 (5.0–13.1)	41.3	20.6	2.0 (1.5–2.6)
Finland									
Round 1/2	14.7	2.5	6.0 (4.9–7.2)	53.4	5.3	10.1 (9.2–11.0)	23.6	11.6	2.0 (1.8–2.3)
Round 2/3	11.8	2.8	4.2 (3.3–5.3)	49.8	4.6	10.7 (9.6–12.0)	27.3	13.3	2.1 (1.8–2.3)
Italy									
Round 1/2	7.0	1.0	6.7 (3.9–11.7)	37.4	3.3	11.2 (8.8–14.2)	31.8	17.2	1.8 (1.6–2.2)
Round 2/3	1.2	0.8	1.6 (0.4–6.6)	25.3	2.6	9.6 (6.8–13.7)	39.3	32.3	1.2 (1.0–1.4)
Netherlands									
Round 1/2	7.2	3.9	1.8 (1.5–2.2)	51.9	13.6	3.8 (3.6–4.1)	26.2	21.0	1.2 (1.2–1.3)
Round 2/3	9.9	2.9	3.4 (2.8–4.3)	55.2	9.2	6.0 (5.5–6.6)	28.2	18.7	1.5 (1.4–1.6)
Sweden									
Round 1/2	14.7	1.0	14.4 (9.8–21.3)	49.1	4.2	11.8 (9.9–14.1)	23.8	15.7	1.5 (1.3–1.8)
Round 2/3	11.7	3.3	3.5 (2.4–5.1)	66.9	10.8	6.2 (5.4–7.2)	15.1	22.5	0.7 (0.5–0.9)
Round 3/4	8.9	3.3	2.7 (1.8–4.1)	58.5	9.4	6.2 (5.2–7.4)	15.4	17.7	0.9 (0.7–1.1)
Round 4/5	10.4	1.5	7.0 (4.6–10.6)	57.8	3.9	15.0 (12.3–18.2)	14.5	10.2	1.4 (1.1–1.8)
Round 5/6	9.9	3.1	3.2 (2.1–4.7)	63.9	7.6	8.4 (7.2–9.9)	17.3	12.0	1.4 (1.1–1.8)

however, age-standardised, indicating that age cannot explain the differences between the centres.

Screening interval could also influence FP result prevalence, as it has been shown that a long screening interval (7 years in the Belgian centre) results in more interval cancers after 4 years from the screen.[26] No difference has been observed in the incidence of interval cancers between the biennial and 4-year intervals in the Swedish and Dutch centres,[10] and according to our results, no conclusive evidence is found to associate FP prevalence with shorter or longer screening intervals. Sweden (with shortest interval) and Belgium (with longest interval) were not the centres with most or least FP results.

Based on our results, the main reason for differences in FP prevalence appears to be PSA threshold which is related to sensitivity. Belgium, the Netherlands and Sweden all used the PSA cut-off of 3.0 ng/ml, while in Finland and Italy a higher PSA cutoff was used. Belgium, Finland and Italy had approximately 10–13% risk for an FP result, which is comparable to a previous estimate from the PLCO trial with a risk of 10% for at least one FP result in annual screening with the PSA cut-off point of 4.0 ng/ml during a 3-year screening period.[6]

This notion is further supported by the observation that if the PSA threshold would have been 4.0 ng/ml in all the centres, the FP prevalence would be 11.7% (instead of 17.8%) and would range from 6.4% (Belgium) to 14.0% (Sweden). The decrease is especially marked in the Netherlands and Sweden. FP prevalence of 11.7% would be relatively similar to the FP risk of 10.4% in the PLCO trial.[6]

The Swedish and Dutch centres had higher frequency of FP results, both exceeding 20%. A previous analysis from the ERSPC study showed that the test sensitivity was slightly higher in the Netherlands (0.93) and Sweden (0.90) compared to Finland (0.89).[27] Of the same five centres analysed in this study, the lead-time for PC has previously been estimated the longest in the Netherlands, possibly reflecting high sensitivity.[28] Based on our results, the enhanced sensitivity (relating to lower PSA threshold, shorter screening intervals) may in turn decrease specificity and cause higher prevalence of FP results. The observed correlation (coefficient 0.78) between PC detection proportion and FP proportion also reflects this.

A further explanation for differences in the FP risk could be some underlying differences in the populations from which the men are selected. It could also be possible that the volunteer-based design of the Dutch centre is subject to selection bias, as e.g. men with BPH underlying lower urinary tract symptoms could be more inclined to participate in a screening study for prostate cancer, yielding the highest FP risk. However, such a bias should also be present in the Belgian results.

Our findings show that men with FP results are very likely to have another FP at re-screening. In fact, more than 50% of the men screened after an FP result still had elevated PSA but negative biopsy. This contradictory result – as 75% men only have one FP result during three rounds – is explained by the fact that many FP men choose not to participate at next screen. We do not know whether these men would be FP or have a PC if they were rescreened. Results from the PLCO trial indicate that men with FP results are almost twice more likely to decline subsequent screening compared to men with a negative screen.[29] Similar results have been shown from the Finnish trial.[5] In this combined analysis of the ERSPC trial, the effect of FP results on non-participation was also found, although weaker.

The PPV of a positive screening test for men without previous biopsy (i.e. new screen-positive men) was similar (22–25%) both at first screen and at later screens, showing that men who turn screen-positive from previous screen-negative results are relatively likely to be diagnosed with PC, consistent with previous studies on the PPV of elevated PSA.[30] This PPV was higher than in men with a previous round FP (14%), but one must bear in mind that from this population, the prevalent PCs have already been "harvested" and that a previous round PSA test cannot be as accurate a predictor as current round PSA test. Based on a single PSA test, the FP men still are at higher risk for next-round PC than screen-negative men. It remains unknown whether this 10% risk for PC in FP men is due to more aggressive follow-up or whether these men carry a genuinely increased risk for PC. The cancers diagnosed after an FP result were mostly localised and low-grade, but 6.3% of the cancers were of aggressive nature (similar to other screen-detected cases).

Our study has some limitations. Variations in the screening protocol in the centres make it difficult to interpret the results, especially if these protocols changed during the screening years (e.g. changes in the PSA threshold or ancillary tests). Variations in the screening protocol in different centres have been previously shown to influence PC detection rates.[31] The strengths of the study include large study size, prospective design and generalizable results.

Our results from a large randomised trial with over 61,000 screened men show that one in six men had at least one FP result during repeated screening protocol. Moreover, almost 20% of men who participated in all screening rounds had FP result(s). The men with FP results are more likely to drop out from subsequent screening but upon attending, they often have another FP result or are diagnosed with PC. The prevalence of FP results could be decreased with higher PSA threshold, but this would also mean missing some cancers. False positive results remain a challenge in screening for prostate cancer and novel approaches are needed to increase the specificity of repeated screening.

Conflict of interest statement

None declared.

Acknowledgements

The international coordination of the European Randomised Study of Screening for Prostate Cancer (ERSPC) has been supported since the study's initiation in 1991 by grants from Europe Against Cancer and the fifth and sixth framework programme of the European Union, by many grants from agencies in the individual participating countries and by unconditional grants from Beckman-Coulter-Hybritech Inc. The studies in each national centre were funded by numerous local grants.

REFERENCES

1. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate cancer mortality in a randomized European study. *N Engl J Med* 2009;**360**:1320–8.
2. Draisma G, Etzioni R, Tsodikov A, et al. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *J Natl Cancer Inst* 2009;**101**:374–83.
3. Etzioni R, Penson DF, Legler JM, et al. Overdiagnosis due to prostate-specific antigen screening: lessons from US prostate cancer incidence trends. *J Natl Cancer Inst* 2002;**94**:981–90.
4. Määttänen L, Hakama M, Tammela TLJ, et al. Specificity of serum prostate-specific antigen determination in the Finnish prostate cancer screening trial. *Br J Cancer* 2007;**96**:56–60.
5. Kilpeläinen TP, Tammela TL, Määttänen L, et al. False-positive screening results in the Finnish prostate cancer screening trial. *Br J Cancer* 2010;**102**:469–74.
6. Croswell JM, Kramer BS, Kreimer AR, et al. Cumulative incidence of false-positive results in repeated, multimodal cancer screening. *Ann Fam Med* 2009;**7**:212–22.
7. Nelen V, Thys G, Neels H, et al. ERSPC: features and preliminary results from the Antwerp study centre. *BJU Int* 2003;**92**(Suppl. 2):17–21.
8. Mäkinen T, Tammela TL, Stenman UH, et al. Second round results of the Finnish population-based prostate cancer screening trial. *Clin Cancer Res* 2004;**10**:2231–6.
9. Ciatto S, Gervasi G, Frullini P, et al. Specific features of the Italian section of the ERSPC. *BJU Int* 2003;**92**(Suppl. 2):30–2.
10. Roobol MJ, Grenabo A, Schröder FH, et al. Interval cancers in prostate cancer screening: comparing 2- and 4-year screening intervals in the European Randomized Study of Screening for Prostate Cancer, Gothenburg and Rotterdam. *J Natl Cancer Inst* 2007;**99**:1296–303.
11. Roobol MJ, Kerkhof M, Schröder 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). *Eur Urol* 2009;**56**:584–91.
12. Mäkinen T, Auvinen A, Hakama M, et al. Acceptability and complications of prostate biopsy in population-based PSA screening versus routine clinical practice: a prospective, controlled study. *Urology* 2002;**60**:846–50.
13. Carlsson S, Holmberg E, Moss S, et al. No excess mortality after prostate biopsy – Results from the European Randomized Study of Screening for Prostate Cancer (ERSPC). *BJU Int* [Epub 2010, Oct 15].
14. Fowler Jr FJ, Barry MJ, Walker-Corkery B, et al. The impact of a suspicious prostate biopsy on patients' psychological, socio-behavioral and medical care outcomes. *J Gen Intern Med* 2006;**21**:715–21.
15. Lafata J, Simpkins J, Lamerato L, et al. The economic impact of false-positive cancer screens. *Cancer Epidemiol Biomarkers Prev* 2004;**13**:2126–32.
16. Finne P, Auvinen A, Määttänen L, et al. Diagnostic value of free prostate-specific antigen among men with a prostate-specific antigen level of <3.0 μg per liter. *Eur Urol* 2008;**54**:362–70.
17. Finne P, Finne R, Bangma C, et al. Algorithms based on prostate-specific antigen (PSA), free PSA, digital rectal examination and prostate volume reduce false-positive PSA results in prostate cancer screening. *Int J Cancer* 2004;**111**:310–5.
18. Schröder FH, Roobol MJ, van der Kwast TH, et al. Does PSA velocity predict prostate cancer in pre-screened populations? *Eur Urol* 2006;**49**:460–5.
19. Etzioni RD, Ankerst DP, Weiss NS, et al. Is prostate-specific antigen velocity useful in early detection of prostate cancer? A critical appraisal of the evidence. *J Natl Cancer Inst* 2007;**99**:1510–5.
20. Bangma CH, van Schaik RH, Blijenberg BG, et al. On the use of prostate-specific antigen for screening of prostate cancer in European Randomised Study for Screening of Prostate Cancer. *Eur J Cancer* 2010;**46**:3109–19.
21. Gupta A, Roobol MJ, Savage CJ, et al. A four-kallikrein panel for the prediction of repeat prostate biopsy: data from the European Randomized Study of Prostate Cancer screening in Rotterdam, Netherlands. *Br J Cancer* 2010;**103**:708–14.
22. Jansen FH, van Schaik RH, Kurstjens J, et al. Prostate-specific antigen (PSA) isoform p2PSA in combination with total PSA and free PSA improves diagnostic accuracy in prostate cancer detection. *Eur Urol* 2010;**57**:921–7.
23. Mehik A, Hellström P, Lukkarinen O, et al. Epidemiology of prostatitis in Finnish men: a population-based cross-sectional study. *BJU Int* 2000;**86**:443–8.
24. Rhodes T, Girman CJ, Jacobsen SJ, et al. Longitudinal prostate growth rates during 5 years in randomly selected community men 40–79 years old. *J Urol* 1999;**161**:1174–9.
25. Wright EJ, Fang J, Metter EJ, et al. Prostate specific antigen predicts the long-term risk of prostate enlargement: results from the Baltimore Longitudinal Study of Aging. *J Urol* 2002;**167**:2484–7.
26. Nelen V, Thys G, Hermans A, et al. Interval cancers in the Antwerp European randomised study of screening for prostate cancer study, using a 6 year screening interval. *Eur J Cancer* 2010;**46**:3090–4.
27. 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* 2009;**18**:2000–5.
28. Finne P, Fallah M, Hakama M, et al. Lead-time in the European randomised study of screening for prostate cancer. *Eur J Cancer* 2010;**46**:3102–8.
29. Ford ME, Havstad SL, Demers R, et al. Effects of false positive prostate cancer screening results on subsequent prostate cancer screening behavior. *Cancer Epidemiol Biomarkers Prev* 2005;**14**:190–4.
30. Roobol MJ, Steyerberg EW, Kranse R, et al. A risk-based strategy improves prostate-specific antigen-driven detection of prostate cancer. *Eur Urol* 2010;**57**:79–85.
31. Otto SJ, Moss SM, Määttänen L, et al. PSA levels and cancer detection rate by centre in the European randomized study of screening for prostate cancer. *Eur J Cancer* 2010;**46**:3053–60.

III

Kilpeläinen TP, Auvinen A, Määttänen L, Kujala P, Ruutu M, Stenman UH, Tammela TL. Results of the three rounds of the Finnish Prostate Cancer Screening Trial – the incidence of advanced cancer is decreased by screening. Int J Cancer 2010;127:1699-705.

uicc
global cancer control

IJC
International Journal of Cancer

Results of the three rounds of the Finnish Prostate Cancer Screening Trial—the incidence of advanced cancer is decreased by screening

Tuomas P. Kilpeläinen[1,2], Anssi Auvinen[2], Liisa Määttänen[3], Paula Kujala[4], Mirja Ruutu[5], Ulf-Håkan Stenman[6] and Teuvo L.J. Tammela[1]

[1] Department of Urology, Tampere University Hospital and University of Tampere, FIN-33521 Tampere, Finland
[2] Tampere School of Public Health, University of Tampere, FIN-33014 Tampere, Finland
[3] Finnish Cancer Registry, Pieni Roobertinkatu 9, FIN-00130 Helsinki, Finland
[4] Department of Pathology, Tampere University Hospital, FIN-33521 Tampere, Finland
[5] Department of Urology, Helsinki University Hospital, FIN-00029, Helsinki, Finland
[6] Department of Clinical Chemistry, Helsinki University Hospital, FIN-00029, Helsinki, Finland

Screening for prostate cancer (PC) remains a controversial issue despite some new evidence on the mortality benefits of PC screening. We conducted a prospective, randomized screening trial in Finland to investigate whether screening decreases PC incidence. Here, we report the incidence results from three screening rounds during a 12-year period. Of the 80,144 men enrolled, 31,866 men were randomized to the screening arm (SA) and invited for screening with prostate-specific antigen test (cut-off 4.0 ng/ml) every 4 years, while the remaining men formed the control arm (CA) that received no interventions. The mean follow-up time for PC incidence in both arms was over 9 years. The incidence rate of PC (including screen-detected and interval cancers as well as cases among nonparticipants) was 9.1 per 1,000 person-years in the SA and 6.2 in the CA, yielding an incidence rate ratio (IRR) 1.5 (95% confidence interval 1.4–1.5). The incidence of advanced PC was 1.1 in the SA and 1.5 in the CA, IRR = 0.7 (0.6–0.8) and the difference emerges after 5–6 years of follow-up. The incidence of localized PC was 7.5 in the SA and 4.6 in the CA, IRR = 1.6 (1.5–1.7). The results from our large population-based trial indicate that screening for PC decreases the incidence of advanced PC. When compared with the CA, the PC detected in the SA there were substantially more often localized, low-grade PCs due to overdiagnosis.

Prostate cancer (PC) is the most common cancer in most industrialized countries.[1] Screening for PC with prostate-specific antigen (PSA) has become a controversial public health issue, and randomized controlled trials are the only reliable way to demonstrate the effectiveness of PC screening. Recently, preliminary mortality results have been published from the European Randomized Study of Screening for Prostate Cancer (ERSPC)[2] and Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO).[3] The ERSPC trial showed a 20% decrease in prostate cancer mortality, whereas in the PLCO trial, no reduction was observed in the screening arm. In both trials, the incidence of PC was substantially higher in the screening arm compared with the control arm, at least partially due to overdiagnosis and lead time bias. It remains to be established whether the benefits of screening for PC with PSA overweigh the harms.[4]

The first analysis of the ERSPC trial indicates a relative reduction in mortality, but there were differences between centers in, *e.g.*, the mode of recruitment, screening interval and biopsy threshold. The early results did not show significant heterogeneity between centers, but the magnitude of effect is likely to differ due to the differences in design and protocol within the study.[2]

We report the results and cancer incidence in screening and control arms in the large population-based Finnish trial during three screening rounds.

Key words: mass screening, prostatic neoplasms, prostate-specific antigen, randomized controlled trials

Abbreviations: CA: control arm; CI: confidence interval; DRE: digital rectal examination; ERSPC: European randomized study of screening for prostate cancer; F/T PSA: free/total PSA ratio; IRR: incidence rate ratio; PC: prostate cancer; PLCO: Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; PSA: prostate-specific antigen; SA: screening arm

Grant sponsor: The Academy of Finland; **Grant number:** 123054; **Grant sponsors:** The Competitive Research Funding (Pirkanmaa Hospital District), Finnish Cancer Organizations
DOI: 10.1002/ijc.25368
History: Received 24 Nov 2009; Accepted 11 Mar 2010; Online 5 Apr 2010

Correspondence to: Tuomas P. Kilpeläinen, Tampere School of Public Health, University of Tampere, FIN-33014 Tampere, Finland, Fax: +358335516057, E-mail: tuomas.kilpelainen@uta.fi

Material and Methods

The Finnish Prostate Cancer Screening Trial is a part of the ERSPC study, which is a randomized multicenter trial. The Finnish trial comprises 80,144 men born in 1929–1944 (aged

55, 59, 63 or 67 years at entry). The subjects were identified from the Finnish Population Registry. After exclusion of men with previous PC diagnosis, a random sample of 8,000 men was allocated to the screening arm (SA) annually in 1996–1999 and the remaining men formed the control arm (CA) that received no intervention.

The men in the SA were invited to a local cancer society clinic for the screening test, *i.e.*, blood sample for determining serum PSA concentration. Men with PSA $\geq$ 4 ng/ml were referred to a local urological clinic for diagnostic examinations including digital rectal examination (DRE), transrectal ultrasound and biopsy. Initially, a sextant biopsy was used, but 10–12 biopsy cores were adopted in 2002. Men with PSA level of 3.0–3.9 ng/ml were referred to an additional test, which in 1996–1998 was DRE and since 1999 a free/total PSA (F/T PSA) ratio with a cut-off point of 16%. Men with a suspicious DRE or F/T PSA ratio $<$ 16% were referred for diagnostic examinations similar to those with PSA $\geq$ 4 ng/ml.

All the laboratory analyses were carried out at the Department of Clinical Chemistry, Helsinki University Hospital. The serum concentrations of total PSA were analyzed by both Hybritech Tandem-E (Beckman Coulter, Brea, CA) and Wallac Delfia (Wallac, Turku, Finland) assays. The free/total PSA ratio was determined with the Wallac ProStatus free/total PSA assay (Wallac).

The men in the SA were reinvited to the second and third screening rounds in a similar manner 4 and 8 years after the first screen (though men older than 71 years of age were no longer invited and thus men aged 67 years at the initial screen were invited only twice). Information on vital status and place of residence was obtained from the Population Register Centre. Men with prostate cancer or emigrated from the study area were not reinvited. These men and those who had died were included in the analyses according to the intention-to-screen protocol. Because of organizational difficulties, there were 1,671 men in the SA who did not receive invitation. These men are included in the analyses as "non-participants" and are analyzed with the SA.

Diagnosis of prostate cancer was based on histopathologic examination, as was determination of the Gleason score. According to the trial protocol, a rebiopsy was indicated if the primary histopathologic diagnosis was prostatic intraepithelial neoplasia, atypical small acinar proliferation or unconfirmed suspicion of prostate carcinoma, or if the PSA concentration was $\geq$ 10 ng/ml. The decision of rebiopsying a patient after a negative biopsy was made by the attending physician, who did not always comply with the protocol of the screening trial. Therefore, some rebiopsies were performed with less strict criteria and some postponed further than protocol-defined time frames.

A screen-detected PC was defined as a cancer diagnosed within 1 year from a positive screening test (due to the fact that not all biopsies or their indications were recorded in the trial database). The PCs that were diagnosed between 1 and 4 years from a positive screening test were categorized separately as early recall PCs (some of these were men who chose to be biopsied at a private clinic or the PC diagnosis was made in a rebiopsy). An interval cancer was defined as a PC within the screening interval in a man with a screen-negative result at the previous screen. A PC with TNM stage of T1-2, N0 and M0 was categorized as localized, whereas a PC with stage T3-4 or N1 or M1 was advanced. An aggressive PC was defined as a PC with one or more of the following characteristics: Gleason score 8–10, T3-4, N1 or M1. The follow-up ended at PC diagnosis, death, emigration or the common closing date (December 31st 2007). Information on cancers detected outside the screening protocol (interval cancers, and those in nonparticipants and the control arm) were obtained from the nationwide, population-based Finnish Cancer Registry, which has 99% coverage of all solid cancers diagnosed in Finland.[5]

The study protocol was approved by Helsinki and Tampere University Hospital Ethical committees. Permission to use cancer registry data was obtained from Research and Development Centre for Welfare and Health (STAKES, currently part of the National Institute of Health and Welfare).

Cumulative incidence in the SA was calculated by dividing the number of PC cases (including all PC cases in the SA) by the number of men in the SA and CA in different screening intervals. Cumulative hazard of PC was estimated using the Nelson-Aalen method.[6,7] Cox regression was used to calculate incidence rate ratios (IRR) and their statistical significance. All statistical analyses were performed using Stata 8.2 (StataCorp, College Station, TX).

Results

In the screening arm, there were altogether 292,474 person-years (pyrs) and in the control arm 449,885 pyrs. The mean follow-up time in the SA was 9.2 years and in the CA 9.3 years (standard deviation in both groups 2.7 years). Because of randomized design, the age distribution in both arms was similar (median 58.7 years at entry in both arms. Age proportions at entry in the SA and CA, respectively, were 55 years: 32.9% *vs.* 33.0%; 59 years: 26.2% *vs.* 26.3%; 63 years: 21.6% *vs.* 21.5%; 67 years: 19.2% *vs.* 19.2%).

In the SA, we invited 30,195 men to the first round (participation proportion 68.8%), 26,240 men (70.9%) to the second and 18,338 men (69.5%) to the third round. A total of 23,771 men (78.7%) participated in at least one screening round, and 10,327 men (52.1% of those invited to all rounds) participated in all the three rounds.

A total 2,655 PCs were detected in the SA and 2,796 PCs in the CA during follow-up (Fig. 1). The total cumulative incidence was 8.3% in the SA and 5.8% in the CA ($p <$ 0.001). The incidence rate of PC was 9.1 in the SA and 6.2/1,000 pyrs in the CA (incidence rate ratio (IRR) 1.46, 95% confidence interval (CI) 1.4–1.5, $p <$ 0.001). Nelson-Aalen cumulative hazard estimates of PC risk were larger for the SA and the difference widened with follow-up (Fig. 2).

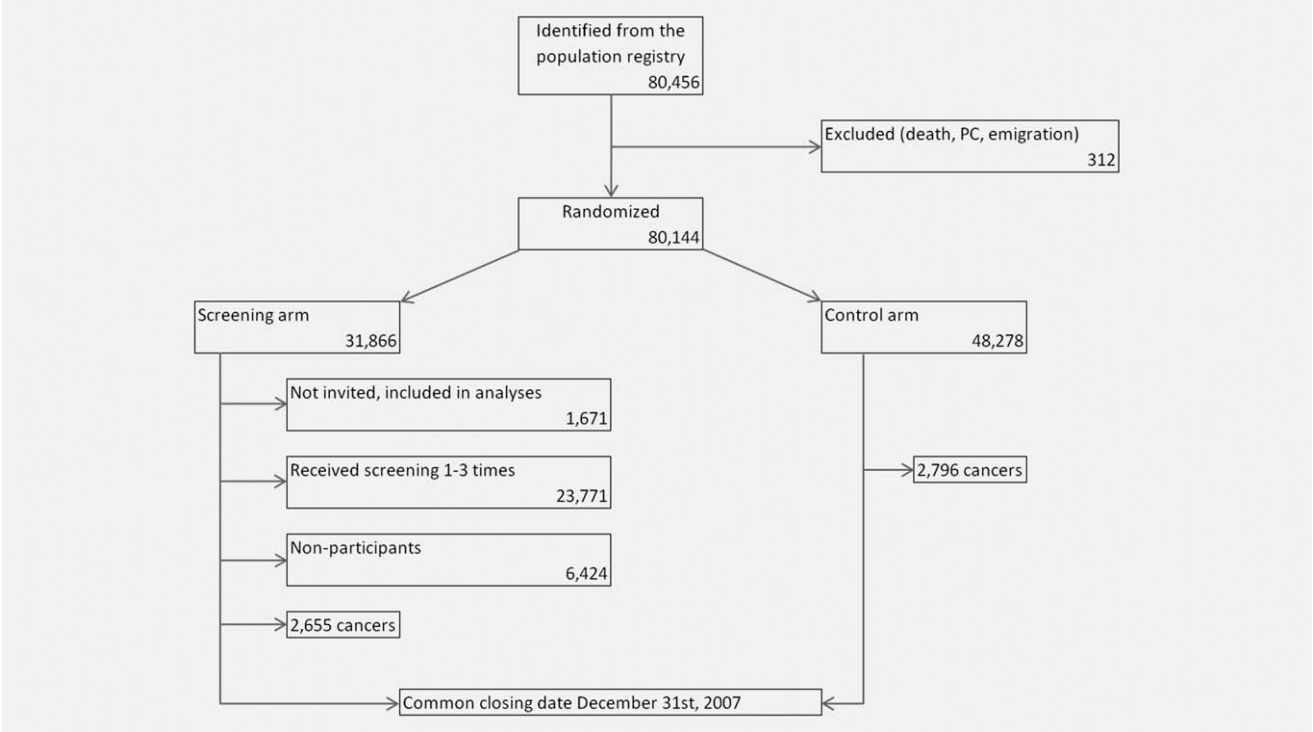


Figure 1. A flow chart of the Finnish Prostate Cancer Screening Trial.

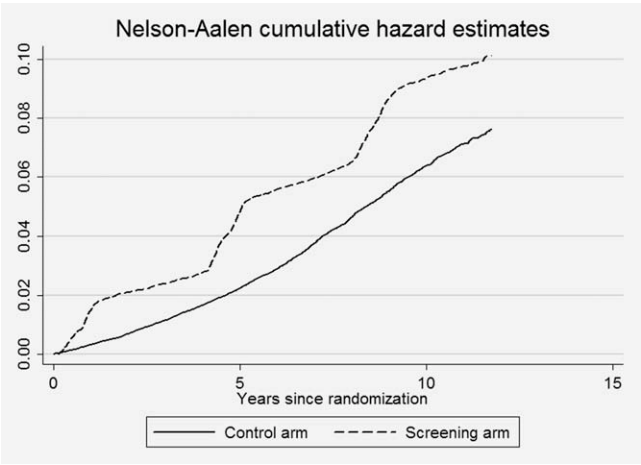


Figure 2. Nelson-Aalen cumulative hazard estimation of overall prostate cancer risk in the control arm and the screening arm.

The proportion of screen-negative men decreased and the proportion of men with screen-detected PC increased with each screening interval (Table 1). The cumulative incidence increased from the first to the second screening interval but decreased in the third screening interval (Table 2) in most age groups and both arms.

The incidence of localized PC was 7.5 in the SA and 4.6/1,000 pyrs in the CA (IRR 1.63, 95% CI 1.5–1.7, $p = < 0.001$). The incidence of advanced PC was 1.1 in the SA and 1.5/1,000 pyrs in the CA (IRR 0.69, 95% CI 0.6–0.8, $p < 0.001$) (Fig. 3). When stratified by age at randomization, the difference was largest in the oldest age group. Among men aged 55 years at entry, IRR was 0.84 (CI 0.6–1.2, $p = 0.33$), in the 59-year-olds, IRR was 0.79 (CI 0.6–1.0, $p = 0.09$), in the 63-year-olds, IRR was 0.83 (CI 0.6–1.1, $p = 0.14$) and in the 67-year-olds, IRR was 0.57 (CI 0.5–0.7, $p < 0.001$). The total cumulative incidence of localized PC was 6.9% in the SA and 4.3% in the CA ($p < 0.001$), and the cumulative incidence of advanced PC was 1.0% in the SA and 1.4% in the CA ($p < 0.001$). In the CA, the proportion of localized PC was lower in all intervals compared with the SA (Table 3). The absolute effect of the reduced incidence of advanced PC can be expressed as number needed to screen, which was 250 (1/(1.4%–1.0%), 95% CI 181–411).

The incidence of low-grade PC (Gleason score 2–6) was 5.8 in the SA and 3.2/1,000 pyrs in the CA (IRR 1.82, 95% CI 1.7–1.9, $p = < 0.001$; cumulative incidence 5.3% *vs.* 3.0%, $p < 0.001$). For Gleason score 7 cancers, the corresponding figures were 1.9 *vs.* 1.8/1,000 pyrs (IRR 1.02, 95% CI 0.9–1.1, $p = 0.72$; cumulative incidence in both groups 1.7%, $p = 0.72$) and for Gleason score 8–10 the incidence was 0.8/1,000 pyrs in the SA and 0.9 in the CA (IRR 0.89, 95% CI 0.8–1.0, $p = 0.16$; cumulative incidence 0.8% *vs.* 0.7%, $p = 0.15$). When only the Gleason 8–10 cancers were analyzed, the cumulative incidence of advanced PC was 0.31% in the SA and 0.50% in the CA ($p = 0.0008$).

Altogether 161 interval cancers were detected, of which 50 (78% localized) after the first screen, 89 (84%) in the second interval and 22 (73%) in the third interval. The cumulative incidence of interval cancer was 0.53% and that of aggressive

Table 1. The results in the three screening intervals

	Interval 1, *N* (%)	Interval 2, *N* (%)	Interval 3, *N* (%)
Screening arm			
Participants			
Screen-negative	18,812 (90.5)	16,309 (87.6)	11,096 (87.1)
False-positive	1,331 (6.4)	1,489 (8.0)	998 (7.8)
Not biopsied	103 (0.5)	202 (1.1)	191 (1.5)
Screen-detected PC	543 (2.6)	613 (3.3)	455 (3.6)
Total	20,789 (100.0)	18,613 (100.0)	12,740 (100.0)
Early recall PC	112	114	26
Interval PC	50	89	22
Nonparticipants	9,406	7,627	5,598
PC	126	157	68
Not invited	1,671	5,626	13,528
PC	50	63	167
Control arm			
PC	757	1,249	790

PC, prostate cancer.

Table 2. The number of prostate cancer cases in both arms by age group and screening interval

	Screening arm			Control arm		
	Interval 1	Interval 2	Interval 3	Interval 1	Interval 2	Interval 3
No. of men	31,866	29,128	25,947	48,278	44,512	39,946
No. of PCs (cumulative incidence, %)						
55 years	125 (1.2)	–	–	97 (0.6)	–	–
59 years	203 (2.4)	233 (2.3)	–	172 (1.4)	260 (1.7)	–
63 years	243 (3.5)	260 (3.4)	240 (2.6)	208 (2.0)	305 (2.6)	196 (1.4)
67 years	310 (5.1)	301 (4.9)	215 (3.1)	280 (3.0)	350 (3.7)	218 (2.0)
71 years	–	242 (4.6)	191 (3.6)	–	334 (4.2)	179 (2.2)
75 years	–	–	92 (2.1)	–	–	197 (3.0)
Total	881 (2.8)	1036 (3.6)	738 (2.8)	757 (1.6)	1249 (2.8)	790 (2.0)

PC, prostate cancer.

interval cancer 0.10%. A total of 176 (46% localized) PCs were diagnosed among either never-participants or previous round nonparticipants in the SA during the first, 220 (63%) in the second and 235 (60%) in the third screening interval (cumulative incidence 8.1%).

Discussion

The results from the Finnish Prostate Cancer Screening Trial show that the overall prostate cancer incidence rate is ~50% higher in the screening than in the control arm during a mean follow-up time exceeding 9 years. This difference is mostly due to the high incidence of low-grade and localized cancers in the SA. Screening succeeded in decreasing the incidence of advanced PC in the SA by a quarter, which is an important intermediate indicator of PC mortality. The absolute effect in terms of reduction in advanced cancer was substantially larger than the reported mortality reduction.[2] This could be due to either lower validity (not all cancers detected earlier due to screening avoid PC death) and also that the lead-time for advanced or aggressive cancers is shorter and is therefore a better indicator of long-term effect.

The Finnish trial is part of the ERSPC study, from which preliminary mortality results were recently reported.[2] This study showed for the first time that screening for PC with PSA can decrease PC mortality with a best estimate of relative risk reduction of 20%. ERSPC is a multicenter study in seven European countries. There are certain differences in the screening protocol between these countries. Therefore, differences among the ERSPC centers in the magnitude of mortality reduction can be anticipated. It is not yet clear how to achieve maximal decrease in PC mortality with minimal harm to the screened population.

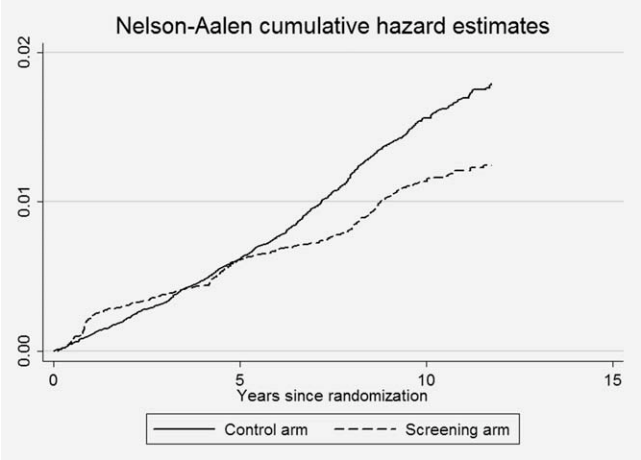


Figure 3. Nelson-Aalen cumulative hazard estimation of advanced prostate cancer (T3-4NxM0 or T1-4NxM1) risk in the control arm and the screening arm.

The main differences in screening protocol between the ERSPC centers are mode of recruitment, screening interval, invitation procedures (*e.g.*, whether to reinvite nonparticipant men or not) and the PSA threshold leading to biopsy. The Finnish trial used a relatively high cut-off level of PSA $\geq$ 4.0 ng/ml and a screening interval of 4 years. Our study was population-based, *i.e.*, based on comprehensive recruitment of all men in the source population to ensure good generalizability and obtain a realistic estimate of the screening effect achievable by mass screening. We reinvited also nonparticipants unless exclusion criteria had been met. Our trial showed adequate participation rate for a population-based study; approximately two thirds of the invited men participated at each of the three rounds.

In both arms, the overall incidence of PC increased initially but leveled off and showed eventually some decline. This temporal pattern (period effect) is parallel to the secular trends in PC incidence in the entire Finland. When men in the SA were compared with men of the same age but with 0–2 previous screens, the cumulative incidence in that interval was the same or somewhat lower in men who had been screened once before, but markedly lower if the men had been screened twice before (*e.g.*, cumulative incidence in 63-year-old men was 3.6% in the first round, 3.5% in the second round and 2.6% in the third round). Also in the CA, when men of same age were compared at different periods, the incidence increased during the second interval and decreased markedly subsequently. One explanation is increasing contamination in the CA, *i.e.*, there was more opportunistic PSA testing in the control men during the second follow-up period compared with the first one. Previously, an overall 20% contamination rate has been estimated in the CA of the ERPSC trial.[8] The frequency of contamination in the control group has not been analyzed in the Finnish trial.

The optimal interval for PC screening is still debated. Some recommend annual screening,[9,10] while the ERSPC study has used a 4-year screening interval with the exception of biennial screening in Sweden. No major differences in the cumulative incidence of (advanced) interval cancers were observed between the Dutch section with a 4-year interval and the Swedish center using biennial interval,[11] suggesting that the longer screening interval was not associated with substantial loss of sensitivity. This is further supported by a recent analysis showing similar test sensitivities in the Dutch (0.95) and Swedish (0.94) ERSPC centers.[12]

Tumor stage and grade provide an intermediate outcome measure of screening efficacy. In a successful screening program, the incidence of advanced and high-grade (hence, less curable) PCs should decrease in the screened group. Our results show a clear stage shift in the SA with a lower overall incidence of advanced PC than in the CA. The Swedish and the Dutch components of the ERSPC study have also demonstrated a favorable stage shift. In Sweden, the cumulative incidence of advanced PC after 8 years of screening with 4

Table 3. Distribution of stage in prostate cancer (PC) categories

	Interval 1, *N* (%)	Interval 2, *N* (%)	Interval 3, *N* (%)
Screening arm			
Screen-detected PC			
Localized[1]	469 (86.4)	577 (94.1)	425 (93.4)
Advanced[2]	74 (13.6)	36 (5.9)	29 (6.4)
n/a	0 (0.0)	0 (0.0)	1 (0.2)
Total	543	613	455
Interval PC			
Localized	44 (88.0)	75 (84.3)	16 (72.7)
Advanced	6 (12.0)	11 (12.4)	6 (27.3)
n/a	0 (0.0)	3 (3.4)	0 (0.0)
Total	50	89	22
Early recall PC			
Localized	105 (93.8)	105 (92.1)	25 (96.2)
Advanced	7 (6.3)	8 (7.0)	1 (3.8)
n/a	0 (0.0)	1 (0.9)	0 (0.0)
Total	112	114	26
Nonparticipant PC			
Localized	80 (45.5)	138 (62.7)	141 (60.0)
Advanced	45 (25.6)	46 (20.9)	40 (17.0)
n/a	51 (29.0)	36 (16.4)	54 (23.0)
Total	176	220	235
Control arm PC			
Localized	536 (70.8)	936 (74.9)	605 (76.6)
Advanced	219 (28.9)	295 (23.6)	170 (21.5)
n/a	2 (0.3)	18 (1.4)	15 (1.9)
Total	757	1249	790

[1]Localized cancer = T1-2, N0 and M0. [2]Advanced cancer = T3-4, N1 or M1.

screening rounds was lower in the SA than in the CA (0.48% *vs.* 0.63%).[13] At the second screen of the Dutch trial, the PC characteristics were more favorable than at the first screen; *e.g.*, the detection rate of stage T3-4 PCs decreased from 18.7% to 3.7%.[14] However, the proportion of advanced PCs may decrease merely by over diagnosing latent PCs—therefore, incidence analyses provide a more valid measure of screening efficacy.

In our study, the incidence of low-grade (Gleason 2–6) cancers was roughly 2-fold in the SA compared with the CA but there was no difference between the SA and the CA in the incidence of Gleason 7 or Gleason 8–10 PCs. The Gleason scores we used in this publication were the original scores, which are subject to changes in Gleason scoring criteria over time.[15] This means that the incidence of low-grade PCs has declined in the recent years most likely because currently similar cancers are assigned a higher grade than before, due to a shift in classification criteria. An analysis taking this bias into account (a random sample from the original biopsies were regraded to match present-day criteria) was recently published from the Finnish trial, showing that the grade of screen-detected cancers was lower compared with that of interval cancers or control arm cancers.[16] Also a joint publication by the ERSPC investigators showed that in Finland a favorable change in PC grades was observed during the first two rounds.[17]

Recently, a study from England[18] showed that PCs detected by elevated PSA are more likely to be less advanced than PCs detected by clinical signs, but no difference was observed in Gleason score 8–10 PCs. This suggests that PSA testing would have no effect on high-grade advanced cancers due to their fast-growing, aggressive nature. In our study, the cumulative incidence of advanced Gleason score 8–10 PC was significantly lower in the SA, indicating that screening could decrease the incidence even in advanced high-grade PCs.

Over diagnosed PCs are latent cancers that are detected because of an intervention, without which the PC would not have been diagnosed during the lifetime of the subject. This is one of the most serious problems with PC screening, as overdiagnosis leads to over treatment, which, in turn, results in adverse treatment effects, psychological stress and increased costs for the health care system.[19,20] The rate of overdiagnosis calculated by different models (for determination of lead-time and it) has been reported to be 23–42%.[19] In the ERSPC mortality analysis, the PC incidence in the SA was 1.4-fold compared with the CA.[2] Our results suggest that ~30% of screened PCs could be over diagnosed in the Finnish trial (calculated as the proportion of excess cases in the SA if the cumulative incidence was the same as in the CA). This is, however, a very crude approximation as overdiagnosis estimations need to be performed with models that take lead-time into account.

There are some limitations to our study. We do not have a reliable estimate of PSA contamination in the control arm but this should cause underestimation rather than overestimation of the observed differences between the SA and CA. For comparability between groups, we used original Gleason scores, which are subject to the change in Gleason scoring criteria over time. Finally, the participation proportion in a population-based study was not as high as can be achieved in a volunteer-based trial (which is biased in other ways). A detailed analysis of nonparticipation is needed to understand why some men choose not to participate. However, our intention-to-screen analysis should not be affected by the nonparticipants.

In conclusion, the reduction in incidence of advanced PC in the Finnish screening study was substantial—the cumulative incidence in the SA was one third lower than that in the CA and the effect was larger than that observed in PC mortality in the ERSPC. The benefits of PC screening are becoming clearer, but more information is needed on the adverse effects, costs and quality of life effects before recommendations on PC screening can be made.

References

1. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin* 2005;55:74–108.
2. Schröder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, Kwiatkowski M, Lujan M, Lilja H, Zappa M, Denis LJ, Recker F, et al. Screening and prostate cancer mortality in a randomized European study. *N Engl J Med* 2009;360:1320–8.
3. Andriole GL, Crawford ED, Grubb RL, III, Buys SS, Chia D, Church TR, Fouad MN, Gelmann EP, Kvale PA, Reding DJ, Weissfeld JL, Yokochi LA, et al. Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med* 2009;360:1310–9.
4. Esserman L, Shieh Y, Thompson I. Rethinking screening for breast cancer and prostate cancer. *JAMA* 2009;302:1685–92.
5. Teppo L, Pukkala E, Lehtonen M. Data quality and quality control of a population-based cancer registry. Experience in Finland. *Acta Oncol* 1994;33:365–9.
6. Aalen OO. Non parametric inference for a family of counting processes. *Ann Stat* 1978;6:701–26.
7. Nelson W. Theory and applications of hazard plotting for censored failure data. *Technometrics* 1972;14:945–66.
8. de Koning HJ, Liem MK, Baan CA, Boer R, Schröder FH, Alexander FE; ERSPC. Prostate cancer mortality reduction by screening: power and time frame with complete enrollment in the European randomised screening for prostate cancer (ERSPC) trial. *Int J Cancer* 2002;98:268–73.
9. Smith RA, Cokkinides V, Brawley OW. Cancer screening in the United States, 2009: a review of current American cancer society guidelines and issues in cancer screening. *CA Cancer J Clin* 2009;59:27–41.
10. American Urological Association Guidelines; Available at: http://www.auanet.org/content/guidelines-and-quality-care/clinical-guidelines.cfm, accessed on July 8th, 2009.
11. Roobol MJ, Grenabo A, Schröder FH, Hugosson J. Interval cancers in prostate cancer screening: comparing 2- and 4-year screening intervals in the European

randomized study of screening for prostate cancer. Gothenburg and. Rotterdam. *J Natl Cancer Inst* 2007;99:1296–303.
12. Auvinen A, Raitanen J, Moss S, de Koning HJ, Hugosson J, Tammela T, Roobol M, Lilja H, Hakama M. Test sensitivity in the European prostate cancer screening trial: results from Finland. Sweden, and the Netherlands. *Cancer Epidemiol Biomarkers Prev* 2009;18:2000–5.
13. 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* 2004;100:1397–405.
14. Postma R, Schröder FH, van Leenders GJ, Hoedemaeker RF, Vis AN, Roobol MJ, van der Kwast TH. Cancer detection and cancer characteristics in the European randomized study of screening for prostate cancer (ERSPC)—section Rotterdam. A comparison of two rounds of screening. *Eur Urol* 2007;52:89–97.
15. Albertsen PC, Hanley JA, Barrows GH, Penson DF, Kowalczyk PD, Sanders MM, Fine J. Prostate cancer and the Will Rogers phenomenon. *J Natl Cancer Inst* 2005;97: 1248–53.
16. Laurila M, Tammela TL, Auvinen A, Isola J, Visakorpi T, Luukkaala T, Määttänen L, Ruutu M, Ala-Opas M, Mildh M, Martikainen P. Biological aggressiveness of prostate cancer in the Finnish screening trial. *Int J Cancer* 2009;124:547–52.
17. van der Kwast TH, Ciatto S, Martikainen PM, Hoedemaeker R, Laurila M, Pihl CG, Hugosson J, Neetens I, Nelen V, Di Lollo S, Roobol MJ, Määttänen L, et al. Detection rates of high-grade prostate cancer during subsequent screening visits. Results of the European randomized screening study for prostate cancer. *Int J Cancer* 2006;118:2538–42.
18. Pashayan N, Pharoah P, Neal DE, Hamdy F, Donovan J, Martin RM, Greenberg D, Duffy SW. Stage shift in PSA-detected prostate cancers—effect modification by Gleason score. *J Med Screen* 2009;16: 98–101.
19. Draisma G, Etzioni R, Tsodikov A, Mariotto A, Wever E, Gulati R, Feuer E, de Koning H. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *J Natl Cancer Inst* 2009;101: 374–83.
20. Etzioni R, Penson DF, Legler JM, di Tommaso D, Boer R, Gann PH, Feuer EJ. Overdiagnosis due to prostate-specific antigen screening: lessons from US prostate cancer incidence trends. *J Natl Cancer Inst* 2002;94:981–90.

IV

Kilpeläinen TP, Tammela TL, Malila N, Hakama M, Santti H, Määttänen L, Stenman UH, Kujala P, Auvinen A. Prostate cancer mortality in the Finnish screening trial. (Submitted)