LIISA MÄÄTTÄNEN

Performance of the Finnish Prostate Cancer Screening Trial Based on Process Indicators

■

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Medicine of the University of Tampere,
for public discussion in the auditorium of Tampere School of
Public Health, Medisiinarinkatu 3, Tampere,
on September 28th, 2007, at 12 o'clock.

UNIVERSITY OF TAMPERE

ACADEMIC DISSERTATION
University of Tampere, Tampere School of Public Health
Doctoral Programs in Public Health (DPPH)
Finnish Cancer Registry
Finland

Supervised by
Professor Matti Hakama
University of Tampere
Professor Anssi Auvinen
University of Tampere

Reviewed by
Professor emeritus Mauri Isokoski
University of Tampere
Professor Olavi Lukkarinen
University of Oulu

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

Tel. +358 3 3551 6055
Fax +358 3 3551 7685
taju@uta.fi
www.uta.fi/taju
http://granum.uta.fi

Cover design by
Juha Siro

Acta Universitatis Tamperensis 1247
ISBN 978-951-44-7026-4 (print)
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 640
ISBN 978-951-44-7027-1 (pdf)
ISSN 1456-954X
http://acta.uta.fi

Tampereen Yliopistopaino Oy – Juvenes Print
Tampere 2007

Contents

List of original publications

This dissertation is based on the following publications, referred to in the text by their Roman numerals

I Määttänen L, Auvinen A, Stenman U-H, Rannikko S, Tammela T, Aro J, Juusela H and Hakama M (1999): European randomized study of prostate cancer screening: first-year results of the Finnish trial. Br J Cancer 79:1210–1214.

II Määttänen L, Auvinen A, Stenman U-H, Tammela T, Aro J, Juusela H, and Hakama M (2001): Three-Years Results of the Finnish Prostate Cancer Screening Trial. J Natl Cancer Inst 93:552–553.

III Auvinen A, Määttänen L, Finne P, Stenman UH, Aro J, Juusela H, Rannikko S, Tammela TL and Hakama M (2004): Test sensitivity of prostate-specific antigen in the Finnish randomised prostate cancer screening trial. Int J Cancer 111:940–943.

IV Määttänen L, Hakama M, Tammela TLJ, Ruutu M, Ala-Opas M, Juusela H, Mildh M, Martikainen P, Stenman U-H and Auvinen A (2007): Specificity of serum prostate-specific antigen determination in the Finnish prostate cancer screening trial. Br J Cancer. 15:56–60.

1 Introduction

The aim of cancer screening is to reduce mortality and to improve quality of life. In general, screening requires that the disease must be common; severe enough to have considerable public health impact and it must have a pre-clinical phase, during which it can be diagnosed and the treatment results are more favourable than at later stages. In addition, a screening method with sufficient sensitivity and specificity must be available and the screening procedure must not cause undue harm to the participants. (Hakama 1991)

Prostate cancer is currently the most common cancer among men in many industrialized countries, including Finland. More than 5,300 new cases are diagnosed annually in Finland and more than 800 men die of prostate cancer each year (www.cancerregistry.fi/eng/, Basic statistics, 4.4.2007).

2 Review of the literature

2.1 Occurrence, etiology and prognosis

Prostate cancer is currently the most common cancer among men in many industrialized countries. In Finland more than 5,300 new cases are diagnosed annually (www.cancerregistry.fi/eng/, Basic statistics, 4.4.2007). In 2002 there were 679,000 new cases of prostate cancer in the world, making this the fifth most common cancer worldwide and the second most common in men (Parkin et al. 2005).

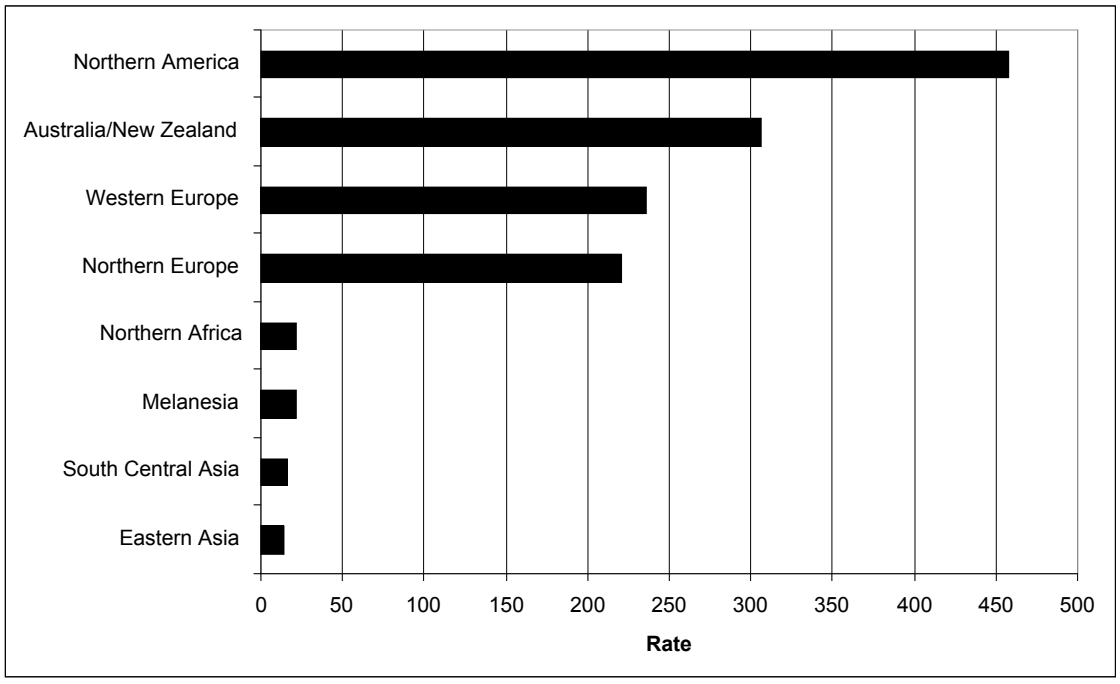

Figure 1. Prostate cancer incidence, at age 45 and above, in eight geographical regions. Rates per 100,000 person-years adjusted to the world standard population (Amended from Globocan 2002).

Prostate cancer incidence differs remarkably between geographical regions. This indicates substantial differences in etiological exposures, but it is also partly due to the variation in the coverage of the cancer registries between regions. (Parkin et al. 2002). Black men in the United States have the highest incidence and mortality from prostate cancer in the world. Increased fat intake may play a role in the greater risk and more aggressive nature of prostate cancers seen in black men. The role of hormones in increasing the risk is still unclear (Freedland et al. 2005). The increasing use of transurethral resection (TURP), a common surgical procedure usually performed to relieve urinary obstruction due to benign enlargement of the prostate, has contributed to the extent of the recorded incidence of prostate cancer (Potosky et al. 1990). Also, the use of prostate specific antigen, PSA, testing in the USA and other industrial countries has led to a dramatic increase in the incidence of prostate cancer since the late 1980s. In the USA incidence rates trends in prostate cancer increased until 1992 and decreased thereafter (Potosky et al. 1995, Legler et al. 1998, Hankey et al. 1999, Fremgen et al. 1999, Ries et al. 2006). Prostate cancer was the second most frequent incident form of cancer in men in Europe with 237,800 new cases estimated in 2004 (Boyle and Ferlay 2005).

Migrants from low-risk countries to areas of higher risk show marked increases in incidence, for example, Japanese migrants living in the United States (Miller et al. 1996).

The etiology of prostate cancer is not yet well understood. Obesity is associated with more advanced disease and worse outcomes in men with prostate cancer. There is limited evidence of associations with obesity and rich fat intake to an increased risk of prostate cancer. (Whittemore et al.1995, Giles and Ireland 1997, Ritch et al. 2007) Findings also suggest that middle-aged men with the metabolic syndrome were more likely to develop prostate cancer (Laukkanen et al. 2004), especially those with low vitamin-D levels (Tuohimaa et al. 2007). A significant trend in the increased risk in men with higher body-mass-index values was observed for death from cancer of the prostate (Calle et al. 2003).

Drinking coffee and alcohol was reported to be unrelated to risk, but there is limited evidence that tobacco may be a risk factor for prostate cancer (Hsing et al 1990, Giles and Ireland 1997, Plaskon et al. 2003, Malila et al. 2006). Long-term supplementation with alpha-tocopherol was shown to substantially reduce prostate

cancer incidence, but abundant use of beta-carotene may contribute to the development of prostate cancer (Virtamo et al. 2003). Low levels of plasma selenium and vitamin-D increase the risk of prostate cancer (Ahonen et al. 2000, Brooks et al. 2001). It is possible that both very low and very high vitamin-D levels increase the risk of prostate cancer (Tuohimaa et al. 2004a, Tuohimaa et al. 2004b, Tuohimaa et al. 2007). Phytoestrogens such as enterolactone are shown to reduce the risk of prostate cancer (Stattin et al. 2002), but the results have been inconsistent (Kilkkinen et al. 2003).

Recurrent or chronic prostate inflammation is likely to have a role in the development of prostate cancer (Nelson et al. 2003). In a five-year follow-up study of 177 patients, chronic inflammation was found to be a significant risk factor for prostatic adenocarcinoma (McLennan et al. 2003). Gland infection causes proliferative inflammatory atrophy (PIA), which is thought to be cancer in situ. The lesions of PIA contain inflammatory cells and they elaborate the epithelial cells and are possibly precursors of prostate intraepithelial neoplasia (PIN) and prostate cancer. High concentrations of cytomegaloviruses were detected in PIN lesions and prostate cancer biopsy specimens (Samanta et al. 2003), but the indication of the virus for the risk of prostate cancer remains unknown. Giles et al. (2003) did not find any association of prostate cancer with the number of sexual partners. There is also recent evidence that prostate cancer may be a viral disease (Klein et al. 2006). It is possible that xenotropic murine-like retrovirus (XMRV) causes mutation and weakens the function of the gene RnaseL (also referred to as the HPC1 gene, i.e., the human prostate cancer 1 gene). The virus was not found in the tissue of the prostate, but in stroma cells that were near the cancer cells. It is possible that it produces inflammation and potential tumour growth.

There is a lack of solid evidence to suggest that infections would increase the risk of prostate cancer. In fact, seropositivity for oncogenic HPV types or for chlamydia trachomatis was inversely associated with prostate cancer (Anttila et al. 2005, Korodi et al. 2005). While inherited genes are much studied, the relatively large effect of heritability in cancer at only a few sites suggests major gaps in our knowledge of the genetics of cancer (Lichtenstein et al. 2000, Matikainen 2001). In prostatic cancer a substantial hereditary component has been proposed (Matikainen 2001). However, in the Finnish component of the ERSPC, the significance of family history was only of marginal importance (Mäkinen et al. 2002b). Because family history indicates both

hereditary risk and a risk of shared environment of the family members, it may be that the hereditary risk is not substantial.

There is some evidence of correlation between testosterone levels and prostate cancer risk (Ross et al. 1986, Parsons et al. 2005) but studies have also reported finding no statistically significant relationship (Dorgan et al. 1998, Stattin et al. 2004). Hormonal replacement therapy is proposed to reduce the risk of prostate cancer (Algarté-Génin et al. 2004), but direct epidemiological evidence is missing. In particular, the effect of vasectomy is a sensitive research subject and the results of increased risk are likely to be biased. Vasectomy was found to increase the risk of prostate cancer in an interview-based study. The survey method used is likely to be affected by bias as patients diagnosed with prostate cancer may feel less embarrassed about sensitive questions on vasectomy than the controls (Sunny 2005).

Various agents with antioxidant, antiproliferative, anti-inflammatory, or pro-apoptotic actions may prevent prostate cancer (Klein et al. 2004). Finasteride is the only agent shown in a randomized trial to decrease the risk of prostate cancer (Thompson et al. 2003) with a 25% effect size point estimate. Even this estimate may be due to a reduction of biopsies in men receiving the medication (Hamdy 2007). In fact, in a population-based case-control study, users of finasteride and alpha-blockers had an increased risk of prostate cancer. A systematic review and meta-analysis on the effect of statins on the risk of prostate cancer resulted in a risk ratio of 1.00 (Browning and Martin 2007).

Results from etiologic and preventive studies of prostate cancer seem to create more questions than they answer (in the form of evidence-based results). Randomized trials could clarify the potential for successful prostate cancer prevention; however, this has not been the case up to now.

Physical activity reduces the risk of cancer (Rintala et al. 2003). There are several studies that consistently show relative risks of less than one (Friedenreich et al. 2004, Giovannucci et al. 2005). The intensity of physical activity and the man's age are related to the risk. Only in a cohort study carried out in the Netherlands (Zeegers et al. 2005) the effect of sedentary work was found to be marginal or nonexistent.

Mortality rates are probably a better indicator of the risk of invasive prostate cancer in different populations than are incidence rates. This is the case even if they depend not only on occurrence, but also on treatment results (Parkin et al. 2002). The

prognosis of prostate cancer is relatively good. In Finland, about 800 men die from prostate cancer each year, but mortality has not changed although the incidence has increased rapidly during the last 50 years. (www.cancerregistry.fi/eng/, Basic statistics, 4.4.2007). Prostate cancer is a relatively minor cause of mortality with 221,000 deaths in the world (5.8% of cancer deaths in men and 3.3% of all cancer deaths) in relation to its incidence, with 11.7% of incident cases in men (Parkin et al. 2005). Mortality rates have risen more rapidly in Asian countries than in high-risk countries, but U.S. blacks still had mortality rates that were 50–60 times higher than the rates in Shanghai, China (Hsing et al. 2000).

In the European Union, prostate cancer ranked as the third most common cause of cancer death in men with 68,200 deaths estimated in 2004 (Boyle and Ferlay 2005). Since the 1990s, mortality rates have declined in several industrialized countries, but it is difficult to evaluate whether this is due to earlier detection or improved treatment (Oliver et al. 2001, Baade et al. 2004, Efstathiou et al. 2006). In the United States, age-adjusted prostate cancer mortality has now dropped below the rate observed in 1986, this fall in mortality has occurred since 1995 for white men and since 1997 for non-whites. This decrease in disease mortality was due to a decline in distant disease incidence, and not to improved survival of patients with distant disease (Chu et al. 2003). Also, the distribution by stage of the tumours changed, in 1992 around 20% of the tumours at the time of diagnosis were metastasized, in 2003 this figure was only around 4% (www.seer.cancer.gov, 2006).

The stage of the disease is usually indicated by the TNM classification system, T1-2 M0 cancers are localized, i.e., confined within the prostate capsule, T3-4 M0 cancers are regional or locally spread and T1-4 M1 cancers have distant metastases. The differentiation of the tumour is graded either by the WHO classification (grades I–III) or more commonly using the Gleason score (Gleason 1992); scores 2 to 6 indicating a high and 8 to 10 a poor degree of differentiation while score 7 is an intermediate group.

The survival rate of patients with T1-2 M0 or Gleason 2–6 cancer is good. On the other hand, M1 or Gleason 8–10 cancers have very poor prognosis. Five-year survival was reported at 84% for localized 65% for regional disease and 25% for patients with distant metastasis in Finland in patients diagnosed during the period 1985–1994 (Dickman et al. 1999).

2.2 Screening for cancer

Screening for cancer has been defined as the identification of unrecognised disease by a relatively simple test (Miller et al. 1991). The objective of cancer screening is to reduce mortality or morbidity and to improve quality of life. However, the primary aim of screening for cancer is to decrease mortality.

General pre-requisites applicable to screening for cancer are widely agreed upon. To be fully justified, these pre-requisites usually require that population-based organized screening should be targeted at a disease, which is common in the population and has public health importance. The target disease must also have a pre-clinical stage during which it can be detected and treated, with a possibility of a favourable prognosis. There should also be a treatment available for the screen-detected disease. In addition, a screening method with sufficient sensitivity and specificity must be available and the screening procedure must not cause undue harm to the participants. The most well-known formulation of criteria was published by Wilson and Jungner in 1968 (Wilson and Jungner 1968).

1. The condition sought should be an important health problem.
2. There should be an accepted treatment for patients with the disease.
3. Facilities for treatment and diagnosis should be available.
4. There should be a recognizable latent or early symptomatic stage.
5. There should be a suitable test or examination.
6. The test should be acceptable to the population.
7. The natural history of the condition should be adequately understood.
8. There should be an agreed policy on whom to treat as patients.
9. The cost should be economically balanced in relation to possible expenditure on medical care as a whole.
10. Case-finding should be a continuing process and not a once and for all project.

A description is offered here for those criteria most relevant to prostate cancer screening. First, prostate cancer is the most common cancer among men in Finland: more than 5,300 cases are diagnosed annually and there are more than 800 deaths due to prostate cancer per year (www.cancerregistry.fi/eng/, 4.4.2007).

Second, it should have the required pre-clinical disease phase, which is detectable and, if the disease is treated at this phase, the outcome of treatment should be improved (Wilson and Jungner 1968). This also implies that the target disease of screening should cause mortality, morbidity or disability that is serious enough to justify the screening programme. Potentially fatal diseases, such as prostate cancer, clearly fulfil this criterion. The criterion of reduced mortality is the first priority for screening serious diseases. So far, it is not known if screening for prostate cancer reduces mortality. Sometimes a screening programme may reduce morbidity or improve the quality of life. However, all screening programmes cause also harm. Non-fatal outcomes are evaluated in screening programmes and the assessment of them is important, when a decision is made whether to screen or not.

Third, it is also required that the treatment will give a more favourable outcome to a screen-detected case than if it would have been found clinically in the absence of screening (Wilson and Jungner 1968). Screening is unnecessary, if disease cases are curable without screening. If there is no treatment for a disease, screening is not justified (Miller et al. 1991). In screening for prostate cancer, it is so far not certain whether or not screening reduces mortality from prostate cancer.

Fourth, the target disease should be common enough to justify the efforts of screening (Wilson and Jungner 1968). The pre-clinical phase is the target of detection and the occurrence of the target disorder is the prevalence of pre-clinical disease in the population to be screened (Cole and Morrison 1978). This frequency is a function of incidence of preclinical and clinical disease, which jointly defines the direction of preclinical phase. A screening programme is a continuous process and involving several screening rounds. The prevalence on initial screening may be substantially different from the prevalence at subsequent screens.

On a more general level, the screening test should be capable of identifying disease at the pre-clinical phase. Supposedly, however, the earliest stages of the pre-clinical phase remain beyond identification for most chronic diseases. The total pre-

clinical phase (TPCP) (Cole and Morrison 1978) starts when the disease begins and ends when cancer surfaces clinically.

The end of the TPCP can be operationally defined, e.g., the time of diagnostic confirmation of the disease in case of no screening. The end of the TPCP would vary widely from patient to patient because of variation in biological growth, in patient and doctor delay, the period between the screening test, and the diagnostic confirmation. The detectable pre-clinical phase (DPCP) is that part of the TPCP during which the disease is detectable by a screening test. The detectable starting point, and thus the length of DPCP or sojourn time (Day and Walter 1984), depends on the screening test used and on the cut-off value of the test.

Screening has harmful effects, which are independent of the effectiveness or beneficial effects of screening (Hakama 1991). The main objective of screening is reduced mortality from the target disease. The treatment of screen-detected disease is often less invasive than for disease detected through other means and causes less adverse effects. Therefore, if the disease is found by screening at a more favourable stage than by clinical symptoms, the need for resources may not be as large as when the disease is detected at a clinically detectable phase. But if disease burden rises due to over-diagnosis, the costs increase as well, although the cost per case might be lowered.

If the screening test is positive, diagnostic examinations are carried out within health care by clinical methods similar to other cases, for cancer usually by histological examination. Cases detected by screening may be abnormal according to routine histopathological criteria, but these changes might not be clinically relevant (behave in a malignant fashion). Some prostate cancers can be classed as latent cancers, because they grow slowly or not at all. Some lesions classified histopathologically as malignant and invasive would not have been found before death, if not screened (Holund 1980, Sakr et al. 1993, Sakr et al. 1994). At present it is not possible to reliably identify which of these lesions would have developed into clinically relevant cancers, if they were not detected and treated while in the DPCP. Overtreating these latent cancers leads to unnecessary morbidity due to the side effects of treatment.

Validity of screening is a summary concept of the success of the process and two indicators define it: sensitivity and specificity. Sensitivity indicates the extent to which the pre-clinical disease is identified and specificity describes the extent to which healthy individuals are correctly identified in the screened population (Miller et al. 1991).

The ability of the screening test to identify accurately the target disease depends on both the target disease and the screening test. Assessment of validity in screening is usually limited to the screening test only (Day and Walter 1984). Test sensitivity indicates how well the test recognizes persons with cancer (Hakama 1991). A person, whose cancer is in the DPCP, should receive a positive test result to be correctly labelled. Other positive test results are false positive. Test specificity indicates how well the test is able to distinguish the disease-free subjects. False positive screening results occur when the test is positive although the tested person has no disease in DPCP. Specificity inversely measures adverse effects of the test. Poor specificity yields false positive results leading to higher costs and other adverse effects.

The result of screening test is not always clearly positive or negative and classification can depend on the subjective judgement of the individual who is interpreting the test. Many factors need to be taken into account, when the cut-off point is decided. Because of subjective and biological variations, the cut-off point of the scale classifying the population in terms of screening positives and negatives can be selected at varying levels. A low cut-off point of the PSA test results in a high sensitivity. In selection of the cut-off value, sensitivity and specificity are inversely related: when sensitivity increases, specificity decreases and, therefore, a screening test has roughly a constant validity. The more completely the true disease is detected, the more false positives will result.

Validity of the test alone is not, however, a sufficient measure of the performance of a screening programme, screening is a public health policy and the validity of the total programme should be assessed (Hakama et al. 2007). Validity of the test is evaluated based on all the lesions detected in relation to disease in DPCP. A positive PSA value is followed by histological examination during the screening episode. The test result and the final diagnosis from confirmatory examinations may be different. Therefore test sensitivity and episode sensitivity should be evaluated separately and represent different characteristics of the screening programme. PSA test has been shown to identify the disease in DPCP (Stenman et al. 1994, Gann et al. 1995, Parkes et al. 1995). However, prostate cancers are diagnosed among screening positive men between the screens, indicating that the confirmation failed to detect the disease. Programme sensitivity is the proportion of the persons diagnosed as a result of the screening programme among the persons with the disease in the target population. Programme

sensitivity also depends on the coverage of the target population by the screening programme and attendance of those covered by the programme. Therefore, the programme sensitivity is usually less than, or equal to, the sensitivity of the screening episode or the test. There may be substantial differences between test sensitivity, episode sensitivity and programme sensitivity, on the one hand, and test specificity, episode specificity and programme specificity on the other (Hakama et al. 2007).

Sensitivity is defined as ability to identify disease in the detectable pre-clinical phase DPCP (Hakama 1991). No method can estimate the sensitivity in a manner which is completely consistent with the theoretical definition. The prevalence of disease in the DPCP is unknown. This is due to the fact that even the perfect test applied to a comprehensive population-sample is bound to miss cases early in the DPCP. Estimation of sensitivity is not straightforward (Day 1985, IARC 2002).

One approach for estimating the sensitivity is based on comparison of clinical cases and controls. This procedure may substantially overestimate the sensitivity, because the cases do not represent preclinical disease, which is the target of screening. When applied to persons with unrecognised disease, the sensitivity of a test is usually much lower than when applied to clinical patients.

A simple indication of sensitivity is the prevalence: incidence ratio, i.e., the detection rate relative to incidence in the absence of screening. Such data give only a crude and often misleading estimate of sensitivity, because the true prevalence is unknown and not all pre-clinical cases would progress to clinical disease (overdiagnosis).

With follow-up and repeated screens of the screened population, new clinical cases appear between screenings and new pre-clinical cases appear at subsequent screenings. These observations are used to estimate sensitivity. An unbiased estimate of sensitivity requires that the expected incidence in the absence of screening is known. Such data are available in an optimal fashion, in randomized trials, from the control population (IARC 2005). For the prostate screening trial, one can record new disease appearing during the screening interval among those with a negative initial screening test. The interval disease is compared with those detected in the control arm, i.e., the incidence of interval disease is related to the incidence of clinical disease in the control population. This incidence method (one minus the ratio of these incidences) gives an estimate of sensitivity. However, de novo disease appears during the interval and not all

the false negatives surface during the follow-up. Therefore, the estimate of sensitivity derived in this way is an imperfect but pragmatic measure that gives comparable estimates in different studies. The estimates of episode sensitivity and programme sensitivity can be obtained in a similar fashion (Hakama et al. 2007).

Specificity has been given relatively little attention in the screening literature. As stated above, depending on the target disease to be screened, specificity may be of major importance. For rare diseases, false negative test results are uncommon. False positives are identified with the diagnostic test following positive screen, given that the subject remains free of disease during the screening interval. Hence, specificity can be estimated in a reasonably accurate manner from the proportion of the true negatives among those free of disease (true negatives and false positives).

Decision-making and risk communication should be based on absolute risks, i.e., probabilities rather than relative risks. It is important to understand the implications, especially for the screenee, of the result of a screening episode. These are given by the predictive values. The positive predictive value is the proportion of those with pre-clinical disease among those with a positive test result. The negative predictive value is the proportion of those who are free of target disease among those with a negative test. The predictive values depend on the validity of the test and on the prevalence of the pre-clinical disease. High predictive values require a valid test and confirmation but, particularly for a rare disease, a positive screening test does not necessarily imply the presence of the disease. In contrast, a negative test gives a very high likelihood of absence of the disease, if the prevalence is low. Many of those who attend a screening programme are in fact seeking reassurance that they do not have the disease (Kauppinen et al. 1970). In practice, this is the most frequent benefit of the screening test for rare diseases.

The length of the DPCP is related to the prognosis: fast-growing cancers, with a short DPCP have a poor prognosis. A screening programme detects a large number of slowly growing cancers when compared with incident disease. Screen-detected disease tends to have a more favourable outcome than clinically detected disease, even in the absence of mortality reduction but because the cancers are biologically different (i.e., more slowly growing). The bias introduced by this selection is called length bias (Feinleib and Zelen 1969). Process indicators of effect are prone to length bias and should not be used to assess effectiveness of screening, i.e., mortality reduction.

The effect of the length bias is maximal for a single screen or for the initial screening of a repeated screening programme. For a programme with repeated screenings, the effect of the length bias is directly related to the screening interval: the shorter the interval, the smaller the bias. Due to a large proportion of fast growing cancers, interval disease usually has a poor outcome. Much can be learned from empirical studies, however. For example, in a Swedish randomized trial (Holmberg et al. 1986), breast cancer diagnosed between mammographic screening examinations did not have a poorer survival compared with other clinically detected cases.

Lead-time (Hutchison and Shapiro 1968) is the amount of time by which the diagnosis of disease is brought forward in time compared with the absence of screening. Results from the screening programmes indicate that the lead-time associated with PSA screening is up to 14 years (Etzioni et al. 1998, Auvinen et al. 2002b, Draisma et al. 2003). By definition, an effective screening programme gives some lead-time and has a maximal lead-time equivalent to the length of the DPCP. Even if screening does not postpone death, survival from the time of diagnosis is longer for a screen-detected case than for a clinically detected case. Valid comparison of survival for screen-detected and clinically detected patients assumes correction for this lead-time bias. Therefore, survival is not valid indicator of screening effect.

Detection of disease by screening may lead to a longer duration of illness and also the side effects of treatment may last longer than if the disease would have been found clinically due to symptoms. Some lead-time is a prerequisite of successful screening but at the same time it may also be a harm of screening because the patient is aware of the disease, with mental adverse effects, for a longer time.

Estimation of length bias and lead time can improve the understanding of the natural history of the disease and the effects of screening (Day and Walter 1984), but these biases make process indicators unsuitable for the evaluation of screening programmes. Impact of screening on the process indicators is a necessary but not a sufficient condition of effective screening.

2.3 Screening tests for prostate cancer

Prostate specific antigen, PSA, is a tumour marker, which was first characterised in 1979 (Wang et al. 1979). PSA is a serine protease produced by prostatic epithelial cells, PSA is abundant in seminal fluid, where its main function is to keep sperm fluid liquid by cleavage of proteins. In prostate cancer patients PSA is elevated 5-10 years before clinical diagnosis (Stenman et al. 1994). In ongoing randomized population based screening studies, serum PSA was found to be elevated above 4.0 ng/ml of 8–12% in those screened (Schröder et al. 1998). In men with serum PSA concentrations of 4.0-10.0 ng/ml, about 20–25% of the men were diagnosed with prostate cancer. Several methods that improve the accuracy of PSA test have been proposed. They include age specific reference values and relating serum PSA to prostate volume, (Oesterling et al. 1995, Mettlin et al. 1996), increase in serum PSA over time (Smith and Catalona 1994) and additional use of the free to total PSA ratio (F/T-PSA) (Stenman et al. 1991).

Estimates of sensitivity for PSA levels $\geq$ 4 ng/ml have ranged from 67% to 90% (Hoffman et al. 2002, Catalona et al. 1991, Brawer et al. 1992, Babaian et al. 1992, Brawer et al. 2000, Catalona et al. 1993, Richie et al. 1993, Mistry and Cable 2003). However, the published studies so far are based on cross-sectional observations without follow-up for interval cancer. Furthermore, most of the studies were subject to verification bias, because men with elevated PSA levels were significantly more likely to be biopsied. This is likely to overestimate the sensitivity (Walter 1999). It seems that the PSA test is valid (Stenman et al. 1994), but that the biopsy often fails to detect the lesion in the DPCP, i.e., it has a poor sensitivity (Norberg et al. 1997, Fink et al. 2003).

In screening for prostate cancer, PSA specificity can be defined as the proportion of the disease-free men correctly classified as prostate cancer-free by the test (PSA value below an agreed cut off value, as 4.0 ng/ml), among men classified as disease-free during the screening episode including both screen-negative men and those who were screen-positive but had a negative biopsy. A method consistent with estimation of sensitivity with interval cancer would take interval cancers into account in the definition of those men who are disease-free. Because interval cancers are few compared to those

classified as false positives by the diagnostic confirmation, the effect remains negligible.

Digital rectal examination (DRE) detects enlargement and other morphological changes (asymmetry, indurations) of the prostate gland. During the examination, a health professional inserts a lubricated, gloved index finger into the rectum. A malignant tumour in the prostate can often be felt as a hard lump. The digital rectal examination is not an exact method of detecting prostate cancer because not all abnormalities in the prostate can be felt through the rectum. DRE has been shown to be an insensitive screening tool for prostate cancer with poor predictive value (Thompson et al. 1987, Chodak et al. 1989). The positive predictive value (PPV) of a suspicious DRE was also found to be low in men with low PSA value (3.0–3.9 ng/ml) and when the DRE was used alone as a screening test (Catalona et al. 1994, Mäkinen et al. 2001).

Transrectal ultrasound (TRUS) provides images of the prostate and surrounding tissue and allows the physician to examine the gland for abnormalities (e.g., hypoechoic lesion) and to assess prostate volume (Rietbergen et al. 1998). In the presence of an abnormal PSA level and/or if the results of TRUS are suspicious for prostate cancer, a prostate biopsy is performed. Overall detection rate for prostate cancer and sensitivity were two times higher for TRUS than for DRE. (Lee et al. 1988) Also the predictive value of DRE and TRUS in men not previously biopsied, varied considerably among the three screening centres in European Randomized Study of Screening for Prostate Cancer (ERSPC) (Roobol et al. 2007). During TRUS-guided biopsy of the prostate, ultrasound is used to help the physician properly place the needle through the rectum to the prostate (Vo et al. 2001). Performance of the DRE and TRUS depend very much on the skills of the clinician (Lan et al. 2007).

2.4 European randomized study of screening for prostate cancer (ERSPC)

In 1990 Schröder and Dennis proposed (Schröder et al. 2003) a randomized study of screening for prostate cancer. Such a study required international collaboration. Plans originated in Belgium and in the Netherlands (Schröder 1993). Finland was the third

partner of the ERSPC (Auvinen et al. 1996). Later on, research groups in France, Italy, Spain, Sweden and Switzerland joined the trial (Schröder et al. 2003). At present, more than 80,000 men have been randomized to the screening arm and over 100,000 men belong to the control arm.

The overall objective of the ERSPC is to estimate the effectiveness and efficacy of screening for prostate cancer with the PSA test. Men aged 50–75 years are screened at least two times with two or four-year intervals. The cut-off level for the test is 4.0 ng/ml. Test-positive men are referred for a prostate biopsy and, if cancer is confirmed, the patient is treated and followed up according to the normal clinical practice. A means of follow-up is established for men in the intervention and control arm to record the incident cases of and deaths from prostate cancer, other deaths and migration. In Finland, this takes place through the Cancer Registry and Statistics Finland. An independent cause of death committee has been established in each country.

Not all the countries follow the same protocol. The target population consists of either volunteers or the general population. The screening interval is two years in Sweden, but four years in other countries. During the trial other minor changes have also taken place in the study protocol.

Each participating country maintains a national database in which information from questionnaire at screening and data from results of screening and follow-up are recorded. This information is compiled into a central ERSPC database maintained in England. The serum samples of PSA test are stored in each of the participating centres.

3 Purpose of this study

The randomized screening trial, European Randomized Study of Screening for Prostate Cancer (ERSPC), is run in eight European countries with the purpose of evaluation of the effect on mortality and quality of life of screening with PSA test. Finland is the largest component of this study. The purpose of this study is to assess the performance of the Finnish prostate cancer screening trial by PSA based on intermediate end-points (process indicators).

The specific aims of this thesis are to analyse:

1) The feasibility of prostate cancer screening in Finland, i.e., participation, and success of implementation of screening. (I, II)
2) The distribution of prostate specific antigen levels (PSA) in a male population. (I, II)
3) The process impact of the programme in terms of detection rate, sensitivity and positive predictive value. (II, III)
4) The potential harm of the programme in terms of specificity. (IV)
5) The biological characteristics of the screen-detected cancers as predictors of final outcome. (II)

4 Material and methods

The Finnish prostate cancer screening trial was started in May 1996 and forms the largest component of the European Randomized Study of Prostate Cancer Screening (ERSPC) (Schröder et al. 2003). The target population of the Finnish trial consists of men born in 1929–1944 and resident in the cities of Helsinki and Tampere with surrounding municipalities of Espoo, Kauniainen and Vantaa, as well as Kangasala, Lempäälä, Nokia, Pirkkala and Ylöjärvi. In the first four years of the study (1996–1999), men born in 1929–1944 were randomized and screened for the first time at the age of 55, 59, 63 or 67 years. The subjects were identified from the Population Register Centre. The only exclusion criterion was a previous diagnosis of prostate cancer. Information on prior prostate cancer was obtained through record linkage with the Finnish Cancer Registry and men with a prevalent prostate cancer were excluded from the study prior to randomization.

A letter of invitation explaining the purpose and procedures of the study was sent to the men in the screening arm. The letter also included information about occurrence of prostate cancer, as well as risk factors and treatment options, including their side effects. In addition, the men were asked to fill in a questionnaire regarding urological symptoms and their treatment, as well as previous PSA tests and family history of prostate cancer. After obtaining a written informed consent, a venous blood sample of 15 ml was drawn in a heparinised Vacutainer tube. After separations, serum was frozen at –20^{o} C and sent to the Helsinki University Central Hospital (HUCH), Department of Clinical Chemistry. PSA determinations were then performed using the Tandem-E assay (Hybritech, BeckmanCoulter, San Diego, CA) and free to total PSA ratios were measured with a Delfia assay (PerkinElmer).

The screening test was based on serum concentration of prostate specific antigen (PSA) with diagnostic examination of all subjects with PSA of 4.0 ng/ml or higher. In order to detect cancers among men with PSA concentrations below 4.0 ng/ml and to improve sensitivity, men with serum PSA concentration 3.0 to 3.9 ng/ml were offered a

supplementary digital rectal examination (DRE) by an urologist. Initially, 119 men with PSA 2.0–2.9 ng/ml were also offered a DRE examination, but this was soon discontinued to avoid loss in specificity. If the DRE finding was suspicious for cancer, the man was referred to a urology clinic for diagnostic examinations. From the beginning of 1999 PSA levels 3.0–3.9 ng/ml with a free to total PSA ratio of <0.16 were defined as positive test result. Men with serum PSA concentrations of 4.0 ng/ml or higher, as well as those with PSA levels 3.0–3.9 ng/ml with a positive result in the supplementary test (suspect DRE or free-total PSA ratio <0.16), were referred to one of the four participating hospitals. Random sextant biopsies as well as DRE and transrectal ultrasound (TRUS) were performed on all referred men, and additional directed biopsies were undertaken if there was a suspicious lesion in DRE, TRUS or both.

If prostate cancer was diagnosed, clinical staging was based on DRE and TRUS, complemented by a bone scan if PSA was above 20.0 ng/ml. Pathological staging and grading of the tumours was carried out in the pathological laboratories of one of the four participating hospitals. Similar treatment protocols were used for both screen-detected cancers and for other patients, for any given age, grade and stage. The treatment options for organ-confined disease included a radical prostatectomy, radical radiotherapy, watchful waiting or hormone therapy. Advanced disease was treated with endocrine therapy (antiandrogen, luteinizing hormone releasing hormone agonist, oestrogen or orchiectomy) or watchful waiting.

The ethical committee in each participating hospital approved the study protocol. Permission for obtaining hospital records was acquired from the Ministry of Social Affairs and Health and for obtaining cancer registry data from the Research and Development Centre for Welfare and Health (STAKES).

The positive predictive value was estimated as number of cancers with positive PSA test among all men with such PSA result. The detection rate was estimated as the prevalence of cancers, i.e., the number of cancers detected at screening among all men screened.

To estimate test sensitivity, interval cancers were defined as cancers detected after a negative screening test (after the screening episode) during the screening interval and excluding cancers detected at the subsequent screen. Test sensitivity was estimated on the basis of incidence of interval cancer in 1996–2000. The start of the follow-up period was defined as the date of screening for the screened men. For non-participants and the

control arm, follow-up started at randomization, i.e., on January 1 each year. End of follow-up was the date of death (from any cause), emigration, prostate cancer diagnosis, time of second screening round for the screening arm and four years from start of follow-up for the control arm or the common closing date (31 December 2000). Interval cancer incidence adjusted for incomplete follow-up (i.e., <4 years) was divided by incidence in the control group corrected for selection. Test sensitivity was estimated as one minus this residual risk proportion (Hakama et al. 2007).

Test specificity was defined as the proportion of the disease-free men correctly classified as negative by the screening test (PSA, with supplementary DRE or free to total PSA ratio determination in the PSA range 3.0–3.9 ng/ml) among men classified as disease-free at the screening episode (including the histological confirmation), and corrected by the interval cases. Therefore, false positives consisted of those with a positive screening test but negative diagnostic confirmation minus interval cancers in these men.

5 Results

In the beginning of each year from 1996 to 1999 of the whole cohort of men born 1929–1944 and residing in the ten municipalities comprising the study area, 8,000 men were annually randomized to the screening group and the remaining men (approximately 12,000) constituted the control group. The whole target population of the study totalled 80,458 men, of whom 32,000 were randomized to the study arm and 48,458 to the control arm. The Helsinki area contributed about three quarters (60,082) of the target population. The number of men invited during the first round was very similar each year. In total 249 men died and 197 moved away from the study area after the randomization and before their invitation letter was sent. Other reasons for not being invited were 'prostate cancer diagnosed after randomization and before the invitation date', 'unknown address' and 'emigration after randomization'. At the time of screening, 30,197 men were eligible in the screening group (Table 1). The number of invitees in 1996, the first screening year, was somewhat smaller, because 247 men did not get the invitation due to an administrative problem.

5.1 Feasibility

In the first screening round, 20,793 (69%) of the 30,197 invited men participated and a blood sample was drawn for determination of PSA. The participation proportion remained relatively constant during the four screening years (Table 1). Participation varied from 65% at age 55 to 72% at age 63 (Table 1). The participation was higher in the Tampere area (76%) than in the Helsinki area (66%) (Table 1). The lower participation rate in the youngest age group appeared mainly in the Helsinki area.

Table 1. Number and proportion of men at first screening round by trial arm, invitation, attendance, year of invitation, age and place of residence in the Finnish prostate cancer screening trial

Background variable	Intervention				Control	Total target population
	Invitees		Attendees			
	Randomized	Eligible at time of invitation*				
	N (%)*	N (%)	N (%)**	(%)***	N (%)*	N
Year of invitation						
1996	8,000 (39)	7,281 (91)	5,050 (63)	(69)	12,398 (61)	20,398
1997	8,000 (42)	7,659 (96)	5,255 (66)	(69)	11,048 (58)	19,048
1998	8,000 (39)	7,653 (96)	5,377 (67)	(70)	12,527 (61)	20,527
1999	8,000 (39)	7,604 (95)	5,111 (64)	(67)	12,485 (61)	20,485
Age						
55	10,495 (40)	9,911 (94)	6,457 (62)	(65)	15,947 (60)	26,442
59	8,378 (40)	7,915 (94)	5,552 (66)	(70)	12,748 (60)	21,126
63	6,931 (40)	6,555 (95)	4,729 (68)	(72)	10,431 (60)	17,362
67	6,196 (40)	5,816 (94)	4,055 (65)	(70)	9,332 (60)	15,528
Area of residence						
Helsinki	24,016 (40)	22,601 (94)	14,989 (62)	(66)	36,066 (60)	60,082
Tampere	7,984 (39)	7,596 (95)	5,804 (73)	(76)	12,392 (61)	20,376
Total	32,000 (40)	30,197 (94)	20,793 (65)	(69)	48,458 (60)	80,458

* Of the target population, ** of those randomized, *** of those eligible

5.2 PSA distribution

Serum PSA concentrations were 4.0 ng/ml or higher in 9% (1,826 men) of the participants, and in 24% (5,068) the concentration was 2.0 ng/ml or higher. The PSA concentration was highest in the oldest age groups indicating a positive correlation between age and PSA level (Table 2).

Table 2. Number and proportion of men at first screening round by serum prostate specific antigen (PSA) concentration (ng/ml), year of invitation, age and area of residence in the Finnish prostate cancer screening trial

Background variable	Serum PSA concentration (ng/ml)											
	0–1.9		2.0–2.9		3.0–3.9		4.0–9.9		≥10		All attendees
	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N
Year of invitation											
1996	3,771	(75)	580	(11)	270	(5)	349	(7)	80	(2)	5,050
1997	3,976	(76)	546	(10)	250	(4)	400	(8)	83	(2)	5,255
1998	4,159	(77)	507	(10)	280	(5)	368	(7)	63	(1)	5,377
1999	3,819	(75)	538	(11)	271	(5)	408	(8)	75	(1)	5,111
Age											
55	5,450	(84)	517	(8)	219	(3)	239	(4)	32	(1)	6,457
59	4,400	(79)	520	(9)	249	(4)	319	(6)	64	(2)	5,552
63	3,338	(71)	548	(11)	300	(6)	453	(10)	90	(2)	4,729
67	2,537	(62)	586	(14)	303	(8)	514	(13)	115	(3)	4,055
Area of residence											
Helsinki	11,287	(75)	1,559	(10)	798	(6)	1,120	(7)	225	(2)	14,989
Tampere	4,438	(76)	612	(11)	273	(4)	405	(7)	76	(2)	5,804
Total	15,725	(76)	2,171	(10)	1,071	(5)	1,525	(7)	301	(2)	20,793

5.3 Positive predictive value and detection rate

Of the 1,826 men with a serum PSA concentration of 4.0 ng/ml or higher, 1,730 (95%) underwent diagnostic examinations, including DRE, TRUS and prostate biopsy. Altogether 542 screening detected cancers were found and the overall PPV was 29% (Table 3). The PPV among men with PSA 10.0 ng/ml or higher was 58%. At the beginning of the trial, 119 men with PSA between 2.0 and 2.9 ng/ml were offered DRE. Of these, seven turned out to be suspicious for cancer and were referred to diagnostic examinations. Among them, three cancers were detected with a positive predictive value (PPV) of 43%. A total of 103 men referred to biopsy were not biopsied for reasons including treatment by another urologist (n = 17), emigration from the study area (n = 6), death (n = 4), refusal (n = 43), or for unknown reasons (n = 33).

Table 3. Number and proportion of men at first screening round by serum prostate specific antigen (PSA), referrals, prostate cancers and positive predictive values (PPV%) in the Finnish prostate cancer screening trial

PSA (ng/ml)	No. of men	No. referred to ancillary test					No. of men biopsied	No. of screen-detected cancers	PPV (%)	
		DRE only	Referral based on			Total			Test*	Episode**
			DRE	F/T	Total PSA					
–1.99	15,725	–	–	–	–	–	–	–	–	–
2 –2.99	2,171	119	7	–	–	119	7	3	0.1	43
3 –3.99	1,071	801	82	60	–	861	135	34	3	25
4 –9.99	1,525	–	–	–	1,525	1,525	1,438	329	22	23
10 –	301	–	–	–	301	301	292	176	58	60
Total	20,793	920	89	60	1,826	2,806	1,872	542	11	29

*among all men within the PSA range
**among those biopsied within the PSA range

We found a PPV for serum PSA concentrations of 4.0 ng/ml or greater of 28%, i.e., 3.5 biopsies per cancer. Lowering the PSA cut-off point to 3.0 ng/ml and abandoning the DRE would have increased referral to diagnostic examinations by 5%, in addition to the 9% of men with a PSA concentration of 4.0 ng/ml or greater.

In all age groups the positive predictive value (PPV) of the screening test varied from 50% to 69% and the episode PPV from 52% to 70%, respectively, in men with PSA value of 10.0 ng/ml or greater. The highest PPV value was observed in the 59 year-old age group.

The overall detection rate of the screening programme was 2.4% among men with PSA $\geq$ 4.0 ng/ml and 2.6% among those with PSA 2.0 ng/ml and above. (Table 5).

Table 4. Number of men at first screening round by age and serum prostate specific antigen (PSA), referrals, biopsies, prostate cancers and positive predictive values (PPV%) in the Finnish prostate cancer screening trial

Age (years) and PSA (ng/ml)	No. of men	No. referred to ancillary test	No. of men biopsied	No. of screen-detected cancers	PPV (%) Test*	PPV (%) Episode**
55 years						
0–1.99	5,450	-	-	-	-	-
2 –2.99	736	35	33	9	1	27
3 –9.99	239	239	221	51	21	23
10 –	32	32	31	16	50	52
59 years						
0–1.99	4,400	-	-	-	-	-
2 –2.99	769	33	32	11	1	34
3 –9.99	319	319	309	75	23	24
10 –	64	64	63	44	69	70
63 years						
0–1.99	3,338	-	-	-	-	-
2 –2.99	848	42	40	9	1	22
3 –9.99	453	453	424	96	21	23
10 –	90	90	86	47	52	55
67 years						
0–1.99	2,537	-	-	-	-	-
2 –2.99	889	39	37	8	1	22
3 –9.99	514	514	484	107	21	22
10 –	115	115	112	69	60	62
Total	20,793	1,975	1,872	542	11	29

*among all men within the PSA range
**among those biopsied within the PSA range

Table 5. Cumulative number and proportion of men in the first screening round by prostate specific antigen (PSA), cumulative number of cancers, cumulative positive predictive value of the test (PPV, %) and detection rate (%) in the Finnish prostate cancer screening trial

Lower limit of PSA	No. of men	No. of cancers	Cumulative PPV of test (%)	Cumulative detection rate (%)
0	20,793	542	2.6	2.6
2	5,068	542	10.7	2.6
3	2,897	539	18.6	2.6
4	1,826	505	27.6	2.4
10	301	176	58.5	0.8

5.4 Sensitivity and specificity

The mean interval cancer incidence by age between the first and second screening rounds was 48/100,000. The mean incidence of prostate cancer in control arm was 394/100,000. The age-adjusted test sensitivity for combination of serum PSA <4.0 ng/ml and ancillary test in range 3.0– 3.9 ng/ml was estimated at 0.85. (Table 6)

The proportion of positive screening findings at the cut-off level of 3.0 ng/ml was 13% and 9% at the cut-off level of 4.0 ng/ml (Table 3). Based on PSA and ancillary tests, 1,968 men were screen-positive and 18,825 negative. Out of the positives, 1,866 (95%) were biopsied during the screening episode and among the negatives 42 interval cancers were diagnosed.

Based on the observations, the specificity of the screening test was estimated as (18,825–42)/(18,825–42+1,358–115), i.e. 0.938 (Table 7). In the first screening round, specificity did vary by age.

Table 6. Number of cancers, person-years at risk and incidence of prostate cancer by age among screen-negative men and in the control group during the first screening interval and follow-up in 1996–2000 in the Finnish prostate cancer screening trial

	Test negative men*			Control group			Sensitivity
Age	Interval cancers	Person-years	Incidence	Interval cancers	Person-years	Incidence	
55	8	16,773	48	70	44,243	158	0.70
59	4	13,836	29	124	34,641	361	0.92
63	7	11,505	61	139	28,299	491	0.88
67	5	9,536	52	206	25,227	566	0.91
Total	24	51,650	48**	539	132,410	394**	0.85***

*PSA <3 ng/ml or PSA 3-4 ng/ml with ether benign DRE finding (1996-1998) or free/total PSA ratio ≥ 0,16 (1999)
**Among screen negative men (I_i), in the control group (I_c)
***Mean over the age groups

Table 7. The number of true negative and false positive screening tests, interval cancers and specificity by age in the first screening round of the Finnish prostate cancer screening trial

Age	Negative test		Positive test, negative confirmation		Specificity
	Total	Interval cancers	Total	Interval cancers	
55	6,153	7	214	20	0.975
59	5,137	11	279	23	0.952
63	4,146	11	409	40	0.924
67	3,389	13	456	32	0.888
Total	18,825	42	1,358	115	0.938

5.5 Biological characteristics of cancers

Of the cancers detected in the first screening round 86% were local and 14% advanced. In the last year of the first screening round (1999), tumours by age were detected in more favourable clinical stage than in previous years. A higher, but non-significant, proportion of cancers was detected in the advanced stage (T3-4) for those in the oldest age group than for those in younger age groups. Tumour stage did not differ between screening areas. In the control group the proportion of organ-confined tumours were smaller than in the screening group (Table 8).

Table 8. Number and proportion of men in the first screening round by clinical tumour stage, year of invitation, age and area of residence in the Finnish prostate cancer screening trial

	Clinical stage			
	T1-2 M0	T3-4 M0	T1-4 M1	Total
	N (%)	N (%)	N (%)	N
Year of invitation				
1996	95 (85)	12 (11)	5 (4)	112
1997	115 (84)	18 (13)	4 (3)	137
1998	109 (85)	14 (11)	5 (4)	128
1999	148 (90)	15 (9)	2 (1)	165
Age (at end of year of randomization)				
55	69 (91)	5 (6)	2 (3)	76
59	112 (86)	15 (12)	3 (2)	130
63	129 (85)	20 (13)	3 (2)	152
67	157 (86)	19 (10)	8 (4)	184
Area of residence				
Helsinki	346 (87)	42 (11)	12 (3)	400
Tampere	121 (85)	17 (12)	4 (3)	142
Total	467 (86)	59 (11)	16 (3)	542
Controls	338 (71)	105 (22)	53 (11)	496*

*missing values N=43

Approximately 80% (433/542) of the prostate cancers detected through screening were well-differentiated (Gleason score <7) (Table 9). This proportion was slightly, but not statistically significantly, smaller in 1999 than during the first three screening years. In Tampere, the distribution of Gleason score was less favourable than in Helsinki ($p< 0.01$). In the control group the distribution of Gleason score was less favourable than in the screening group, but number of missing values in the control group was noticeable. (Table 9)

Table 9. Number and proportion of men in the first screening round by Gleason score, year of invitation, age and area of residence in the Finnish prostate cancer screening trial

Background variable	Gleason score				
	2–6	7	8–10	Unknown	Total
	N (%)	N (%)	N (%)	N (%)	N
Year of invitation					
1996	91 (81)	13 (12)	6 (5)	2 (2)	112
1997	117 (85)	15 (11)	4 (3)	1 (1)	137
1998	102 (80)	17 (13)	8 (6)	1 (1)	128
1999	123 (75)	26 (16)	14 (8)	2 (1)	165
Age (at the end of year of randomization)					
55	65 (86)	6 (8)	4 (5)	1 (1)	76
59	109 (84)	13 (10)	8 (6)	- (-)	130
63	121 (80)	17 (11)	11 (7)	3 (2)	152
67	138 (75)	35 (19)	9 (5)	2 (1)	184
Area of residence					
Helsinki	338 (84)	45 (11)	15 (4)	2 (1)	400
Tampere	95 (67)	26 (18)	17 (12)	4 (3)	142
Total	433 (80)	71 (13)	32 (6)	6 (1)	542
Controls	388 (71)	93 (19)	44 (8)	64 (9)	539

6 Discussion

The Finnish prostate cancer screening trial is population-based, i.e., we were able to define the study base and identify all subjects in it. Compared with volunteer-based studies, the population-based approach has the advantage of permitting estimation of effects in the general population, i.e., screening implemented as public health policy. We had a high participation rate (69%), which is of prime importance in achieving representativeness of the population. The somewhat lower participation in the Helsinki area is to be expected given the degree of urbanisation and is similar to participation in cervix and mammographic screening. (Kallio et al. 1994, Anttila et al. 2002). Demographics and health care services in the Helsinki area differ from the rest of the country and even if information on these is not essential for effectiveness estimates, it is useful in assessing of the process indicators. Information on prostate cancer incidence and mortality in the control group was obtained through record-linkage (Finnish Cancer Registry and Statistics Finland). Prostate cancer among men from both the screening and the control arm were treated almost exclusively at the participating clinics. In the screening arm, two of the radical prostatectomies were undertaken at a private clinic.

The Finnish trial is designed as a low-intensity intervention: the cut-off level for PSA is higher and screening interval longer than in most other studies. This approach was chosen to minimize both the costs and the potential adverse effects. The relatively low cost of undertaking this screening programme is likely to increase the probability of it being considered cost-effective, given that mortality reduction and quality of life effects compare favourably with more intensive screening regimens. The overall detection rate of prostate cancer during the first screening round in the Finnish trial (approximately 2%) is somewhat less than indicated by the previous results from studies using a similar screening algorithm (Catalona et al. 1991). Studies using other screening modalities combined with PSA have had higher detection rates (Mettlin et al. 1996, Schröder et al. 1996, Thompson et al. 2006). In early reports, lower detection rates have been reported for subsequent screens following the first round (Labrie et al. 1996, Smith

et al. 1996), but also increase in the detection rates by screening interval were observed (Nelen et al. 2003, Hugosson et al. 2004).

In the early years of the project (1996–2001) sextant biopsy was the standard in the trial and in clinical practice. In autumn 2002 the practice was changed to a 12 core biopsy, because of similar change in the clinical practice. Such a change was likely to improve episode sensitivity (Eichler et al. 2006, Elabbady and Khedr 2006, Emiliozzi et al. 2004, Singh et al. 2004, Siu et al. 2005), but it was not optimal from the point of view of time trends. However, comparability between the arms was maintained, which is the prime objective to maintain validity and to evaluate ultimate effectiveness without bias. All biopsies in this report are sextant ones as the intake was closed in 1999.

The Gleason score is an indicator of the aggressiveness of the tumour, scores 2–6 indicating a slow growing tumour, 8–10 aggressive ones. The intermediate score of 7 was proposed to be divided into two components (Tollefson et al. 2006) depending on the largest component of the tumour. In this study, the original classification used from 1990's in the ERSPC was used to maintain comparability with other ERSPC study centres.

Not all men complied with random allocation – in addition to non-attenders in the screening arm, there is likely to be contamination (opportunistic screening) in the control arm. Information on contamination was not available systematically, but 7–14% of the men entering the screening arm of the trial reported a prior PSA test. Contamination reduces the exposure contrast and therefore decreases the statistical power to detect a mortality difference. Yet, it does not bias the results, as the study is designed as an effectiveness trial, with results indicating the amount of benefit could be achieved by providing organised screening. In such context, the relevant comparison is the realistic context with the level of screening that the population receives, instead of completely unscreened population. Also, some men in the screening arm were not invited due to errors in administrative procedures. This is however unlikely to affect the results, as the proportion of such men was minimal (<1%).

Ideally, a screening test should be able to classify correctly both subjects with and without the target disorder. In practice, the distributions of test values between these two populations always overlap. Validity is the capacity to distinguish normal and high-risk persons. Sensitivity indicates the success in identifying affected subjects. Specificity, on the other hand, shows the potential to correctly identify those free of the

disorder. Specificity is a characteristic of the screening test and an indicator of performance used in evaluation of screening, but not directly applicable in decision-making at individual level.

Specificity and sensitivity are complementary characteristics in the sense that there is a trade-off involved in deciding the threshold: gain in one is balanced by loss in the other. In this study sensitivity was estimated as 85% and specificity 94% and the balance between these two elements depends on the relative impact of false positive versus false negative findings. In the context of prostate cancer, the question is how many negative (unnecessary) biopsies one is willing to accept in order to detect (or miss) one case of prostate cancer. Further, histological findings consistent with prostate cancer do not accurately predict prognosis. Prostate cancer commonly has an indolent disease course (natural history) and a slow progression rate. Therefore, benefit from treatment may be small and over-treatment presents a challenge for screening, as death from other causes will frequently occur before the potential death from prostate cancer.

The detection rate with digital rectal examination among men with PSA below 4.0 ng/ml was an order of magnitude lower than among men with PSA levels above 4.0 ng/ml. The detection rate for men with PSA in the range 2.0–3.9 ng/ml (3.1%) was close to that of the whole study population regardless of PSA concentration (2.3%), which suggests that men with this PSA level do not have a materially increased risk of prostate cancer compared with the general population. The ERSPC trial protocol is based on total PSA concentration, but incorporation of the free to total PSA ratio could decrease the number of false positive test results (Stenman et al. 1994) and improve the specificity of the screening protocol. Finne et al. found that the AUC (area under the curve) of receive operating characteristic improved from 0.55 to 0.73 with additional information from the ratio of free to total PSA compared to PSA only in sera with PSA concentrations of 4.0–10.0 ng/ml. Instead, the ancillary examinations had only a small effect on sensitivity but decreased specificity (Finne et al. 2000).

The positive predictive value of PSA test with cut-off level of 4.0 ng/ml was 25%, which is comparable with mammography screening (Hakama et al. 1991, United Kingdom Trial 1992, Kerlikowske et al. 1993). It is possible that the positive predictive value could be further improved by determining the free to total PSA concentration in addition to the total PSA (Stenman et al. 1991, Lilja et al. 1991).

We defined test sensitivity as the proportion of men with a positive test among all men with prostate cancer in the detectable preclinical phase. As men with preclinical disease among screen-negative subjects cannot be identified, interval cancers (cases surfacing clinically during the first screening interval) were used as an indicator for cancers missed at screening.

Because of over-diagnosis of prostate cancer at screening, we estimated sensitivity based on proportional incidence by the incidence method, i.e., the ratio of interval cancer incidence in screening group relative to cancer incidence in the control arm. (Day 1985, IARC 2002, Ciatto et al. 1995) Generally, test sensitivity is calculated by comparison between the screened men and the control arm. This comparison is asymmetrical in the sense that non-participants are excluded from the screening arm. Correction for the selection effect of attendees in the screening arm was not carried out because the negligible effect as the control population and the non-attenders had practically the same risk.

Our estimate of test sensitivity was 0.85. The interval cancer incidence remained well below the rates in the control arm throughout the screening interval. Strict definition of interval cancer is difficult. A positive screening test was in some cases followed by repeated biopsies, which was not in accordance with the trial protocol. A number of cases were detected due to a positive screening test, but classified as interval cancers. Taking into account overdiagnosis during the first year of follow-up did not substantially affect sensitivity. These results are consistent with serum bank studies before the PSA era showing that serum PSA levels are elevated at least five to six years prior to clinical diagnosis of prostate cancer (Stenman et al. 1994, Hakama et al. 2001, Carter et al. 1992, Gann et al. 1995). These findings suggest that acceptable sensitivity could be achieved even with a re-screening interval longer than four years, which is consistent with the results on lead-time (Auvinen et al. 2002b, Draisma et al. 2003). Furthermore, intermediate PSA levels (3.0–4.0 ng/ml) contributed only marginally to the sensitivity (8% of interval cancers), suggesting that intermediate PSA concentrations are not critical for the effectiveness of screening. Therefore, it seems that there is a window of curability of PSA levels that exclude both high (incurable) and low (insensitive) values.

The specificity of a screening test is an indicator of the adverse effects of screening, including the cost. Specificity determines the some of the costs and the acceptability of a prostate cancer screening programme. To increase specificity, we used a screening algorithm with a relatively high PSA cut-off point and few auxiliary interventions. Specificity and PPV increased rapidly with increasing concentrations of serum PSA, indicating less harm and lower costs per screen-detected cancer at higher cut-off levels. Specificity and the related PPV are important characteristics of a screening test, because screening always has negative effects and false positive results represent the most frequent untoward consequence of screening in most settings, particularly when the target condition is rare. The harms of screening in false positive cases include cost and inconvenience from diagnostic intervention, as well as fear and anxiety from being labelled as having high risk of disease. In prostate cancer, this involves digital rectal examination, transrectal ultrasound and prostate biopsy. The first two may be unpleasant but biopsy carries also some risk, e.g., septic infection, as well as less severe effects such as haematuria and haematospermia. This can result in stigmatization and impaired perceived health. Yet, the short-term quality of life effects of cancer screening are relatively minor (Absetz et al. 2003, Essink-Bot et al. 1998, Brett et al. 1998). Lower participation at subsequent screening has been reported among men positive finding classified as false positive in the first screening round (Mäkinen et al. 2002a).

The mortality reduction through screening is achieved by detecting and effectively treating cases that are potentially lethal, but are detected at a curable stage by screening. In our study, approximately nine out of ten screen-detected cancers were well or moderately differentiated and organ-confined. This is comparable with other studies (Labrie et al. 1996, Mettlin et al. 1996, Smith et al. 1996, Schröder et al. 1996). Screen-detected cancers were detected substantially more frequently at a curable stage, i.e., organ-confined, than clinically detected prostate cancers. Typically, a larger proportion of well-differentiated, slowly growing tumours are detected in the first (prevalence) screening round than subsequent (incidence) screening rounds, because of length bias (Hakama 1991). Improvement of stage and grade distribution compared with otherwise detected cancers is a necessary, but not sufficient proof of effectiveness of a screening programme. Furthermore, the improvement should be measured by the rate, not by the proportion, of advanced or aggressive disease. (Carter et al. 1992, Epstein et

al. 1998, IARC 2002) The detection rate for Gleason grade 7–10 cancer was 0.5%, which indicates that the screening was able to detect not only indolent but also aggressive cancers.

Prostate cancer is currently the most common cancer among men after skin cancer in several industrialised countries (Parkin et al. 2002). Because of its poorly understood etiology, primary prevention is not currently feasible and, therefore, there is considerable interest in screening as a potential approach to prostate cancer control. The only means to establish the effect of screening is to conduct large randomized controlled trials with both mortality and quality of life as end-points (Denis et al. 1995, Auvinen et al. 2002a**,** Auvinen and Hugosson 2003). The definitive evaluation of effectiveness of a screening trial is based on mortality reduction. However, in the case of prostate cancer, the possible mortality reduction may not be manifest prior to at least 10–15 years of follow-up. Preliminary evaluation can be based on intermediate end-points including feasibility of screening, distribution of prostate-specific antigen in population, validity of the programme and biological characteristics of the screen-detected cancers. In general, a screening programme cannot be evaluated on the basis of intermediate end-points alone.

The results of the process indicators of the Finnish prostate cancer screening trial show that screening is acceptable for the target population. The performance of the screening test in terms of specificity and sensitivity is adequate, and the detection rate of aggressive, potentially lethal cancer is reasonable. These results pertaining to intermediate indicators do not provide enough indication for the effectiveness of prostate cancer screening as a public health policy. Before prostate cancer screening can be accepted as a public health policy, a reduction in mortality should be demonstrated and the quality of life effects and cost-effectiveness evaluated. The process indicators show that continuation of the Finnish prostate cancer screening trial is justified.

7 Summary

The aim of cancer screening is to reduce mortality and improve quality of life. Evaluation of this effectiveness ultimately requires extensive follow-up. Therefore, intermediate or process indicators are used as necessary, but not sufficient, early evidence with the potential to predict the mortality effect. For prostate cancer, the effectiveness of screening is unknown and two large-scale trials are being carried out.

The Finnish prostate cancer screening trial is the largest component of the European Randomized Study of Screening for Prostate Cancer. The purpose of this thesis is to assess the Finnish prostate cancer screening trial with PSA, using intermediate endpoints, i.e., participation, PSA distribution, detection rate, sensitivity, specificity, predictive values and prognostic factors of screen-detected cancers.

Altogether 30,197 men were invited and 20,793 (69%) attended. Among attenders 1,826 (9%) were screen-positives with PSA $\geq$ 4.0 ng/ml and additional 1,075 men were referred for diagnostic examination based on DRE or free to total PSA ratio among men with moderately increased level of PSA (3.0–3.9 ng/ml).

Altogether 542 cancers were detected with detection rate of 2.6%. The positive predictive value of the PSA test at cut-off level of 4.0 ng/ml was 28%.

The sensitivity by age of the PSA test at PSA 4.0 ng/ml and with ancillary tests of TRUS and free to total PSA ratio determination was 85% during the 4-year screening interval. Specificity at the cut-off limits of the PSA test at PSA 3.0 ng/ml and with ancillary tests of DRE, TRUS and free to total PSA ratio was 94%.

The clinical stage distribution of the screen-detected cancers was favourable (T1-T2 M0) in 86% and only 3% of the cancers had distant metastases (M1). Of the screen-detected cancers, 6% were aggressive (Gleason 8–10).

The Finnish prostate cancer screening trial is acceptable to the target population, the test is capable of identifying a small proportion of men with a high risk of prostate cancer. Detection of aggressive, potentially lethal cancer implies that PSA testing does not detect only indolent cancers. The results on the process indicators justify the continuation of the Finnish prostate cancer screening trial for estimation of the ultimate effectiveness in terms of mortality reduction.

8 Kiitokset

Tämä työ on tehty Suomen Syöpärekisterissä vuosien 1996 ja 2007 välisenä aikana osana eurooppalaista yhteistutkimusta European Randomized Study of Prostate Cancer Screening (ERSPC).

Seulontakoe ja väitöskirjatyö ovat niin kiinteästi yhteen nivoutuneita, ettei kiitoksia voi erottaa koskemaan vain toista aluetta.

Ohjaajaani professori Matti Hakamaa kiitän hänen sitkeästä ja peräänantamattomasta jääräpäisyydestään. Ilman hänen päättäväisyyttään ei tutkimusta olisi Suomessa voitu aloittaa. Korvaamaton apu on ollut hänen lyömätön ja laaja tietämyksensä seulonnoista ja epidemiologiasta. Hän on myös väsymättä paimentanut väitöskirjatyötä eteenpäin.

Suuret kiitokset kuuluvat toiselle ohjaajalleni professori Anssi Auviselle. Hänen energinen ja positiivinen otteensa kaikkeen on tuottanut tulosta myös tämän työn osalta ja nuo alati esiin pulpahtavat ideat ovat pitäneet vauhtia yllä. Hänen osuutensa artikkelien tuottamisessa on ollut ratkaisevaa. Anssi on monesti mahdollistanut projektin jatkumisen rustaamalla lukemattoman määrän apurahahakemuksia.

Syöpärekisterin Joukkotarkastusrekisteri on kokenut paljon muutoksia näinä vuosina, mutta tämän projektin alusta alkaen on tutkimussihteeri Minna Heikkilä tukenut ja tuuppinut, puolustanut ja auttanut eteenpäin. Minnalta olen oppinut sen minkä tiedän seulontojen käytännön organisoinnista ja hänen sihteerintaidoistaan olen yrittänyt omaksua edes murto-osan. Minnan kanssa käydyt pohdinnat ja keskustelut ovat olleet hedelmällisiä ja ne tietenkin jatkuvat yhä.

Toimiessaan Suomen Syöpärekisterin johtajana emeritusprofessori Lyly Teppo teki tutkimuksen kannalta ratkaisevan päätöksen antamalla tutkimukselle aloitusluvan ja kotipaikan Syöpärekisterissä. Täten hän mahdollisti seulontatutkimuksen alkamisen.

Professori Teuvo Tammelalla on ollut painava osuus tutkimuksen urologian asiantuntijana, henkisenä kannattelijana ja käytännön esteiden poistajana. Hän on arvovallallaan käynyt käsiksi ongelmiin, eivätkä näistä vähäisempiä ole olleet taloudelliset

vastukset. Teuvon hämmästyttävän ehtymätön energia on ulottunut tutkimuksen kaikille alueille ja hän on positiivisella ja valoisalla asenteellaan kannustanut myös tämän opinnäytteen etenemisessä.

Professori Ulf-Håkan Stenmania kiitän toimimisesta väitöskirjan seurantaryhmässä yhdessä professori Teuvo Tammelan kanssa. Tämän lisäksi professori Stenman on eri yhteyksissä jakanut niin paljon tietoa kliinisen kemian alaan liittyvistä aiheista ja PSAsta, etten usko siitä pystyneeni omaksumaan kuin murto-osan.

Suomen Syöpärekisterin nykyinen johtaja professori Timo Hakulinen on ollut tutkimuksen vankka tukija ja ajoittain suorastaan pelastanut projektin jatkumisen jämäkkyydellään. Häneltä on myös riittänyt huomiota ja tukea väitöskirjatyön etenemiselle, johon toimi Kansanterveyden tutkijakoulun johtajana antoi lisäpontta.

Eräältä Paradox-kurssilta löytyi ATK-spesialisti Jarkko Koivuniemi DataFrank oystä. Häneltä lähiverkon toteutukset, tietokantaohjelmien sovellukset ja ohjelmoinnit sekä koko tietokannan rakenteiden kehittely kävivät leikiten ja ratkaisut ongelmiin ja toivomuksiin toteutuivat aina. Jarkon ansiosta PSA-tulokset saatiin oikeille paikoilleen ja vastauskirjeet tuotettua miehille. Samalla hän tuli opettaneeksi ohjelman käytön ja hallintaakin jutustelun höystäessä istuntoja.

PSA-näytteet otetaan Helsingissä Bulevardin Erikoislääkärit, Syöpäjärjestöjen poliklinikalla ja Tampereella Pirkanmaan Syöpäyhdistyksessä. Näiden laboratorioiden henkilökuntaa haluan kiittää joustavasta yhteistyöstä, miesten ohjauksesta ja tutkimuslomakkeiden keräämisestä ja perille huolehtimisesta.

PSA-määritykset tehtiin aluksi HYKSin Kliinisen kemian laboratoriossa ja nykyään Biomedicum Helsingissä, professori Ulf-Håkan Stenmanin johdolla. Muutosta huolimatta työntekijät ovat pysyneet samoina ja olenkin suuren kiitoksen velkaa laborantti Maarit Leinimaalle ja laboratoriohoitaja Anne Ahmanheimolle, jotka taitavasti ovat huolehtineet näytteistä ja selviytyneet työstään kaikenlaisten, eikä vähiten teknisten ongelmien keskellä.

Tietokannan sisällön edellytyksenä on sairaaloista saatu tutkimustieto. Haluan kiittää tutkimushoitajia Helsingin Meilahden ja Marian sairaaloissa, Espoossa Jorvin sairaalassa, Vantaalla Peijaksen sairaalassa ja Tampereen yliopistollisessa keskussairaalassa. He ovat jakaneet lomakkeita, valvoneet niiden täyttämistä ja toimittaneet ne perille.

Tietoja ovat tallentaneet Joukkotarkastusrekisterissä vuosien mittaan useammatkin nuoret ja iäkkäämmät henkilöt lyhyemmän tai pidemmän jakson. Kaikki on kuitenkin tallessa ja siitä heille kiitos.

Kiitän esitarkastajiani professori Olavi Lukkarista ja emeritusprofessori Mauri Isokoskea heidän työstäni antamista kommenteista ja tuiki tarpeellisten täydennysten neuvomisesta.

Koko PSA-tutkimusryhmä, joka on laajentunut toiminnan vuosien karttuessa, on ollut monialainen tiedon tausta. Erityisesti haluan kiittää ylilääkäri Paula Martikaista, koska hänen kanssaan tehty yhteistyö on opettanut paljon ahkeruudesta ja patologiasta.

Koko Joukkotarkastusrekisterin henkilökuntaa, joka kasvaa ja vahvistuu, sietää kiittää paitsi työhön saamastani tuesta niin varsinkin niistä taukojen ja vapaa-ajan riemuista, joita en voi enkä halua unohtaa. Kahvitaukojemme naurunremakka on todennäköisesti jo nyt pidentänyt meidän ikäämme huomattavasti ja jos meno jatkuu, niin nuoruuden lähde on löytynyt.

TietoEnatorissa (nykyään 121 Oyj) on kaikki nämä vuodet Antti Hossa hoitanut kutsutietojen päivityksen, joten kutsut on saatu miehille, joille ne kuuluivatkin. Antin sitoutuminen projektiin on ollut huolehtivaista ja myös erikoistilanteissa hän on pystynyt järjestelemään työnsä meidän eduksemme.

Toimistosihteeri Marita Hallille kuuluu suuri kiitos, sillä hän auttavaisen luontonsa lisäksi omaa paljon kokemusta väitöskirjaoppilaiden luotsauksessa läpi käytännön karikoiden ja tästä olen minäkin saanut osuuteni, jota ilman ei päätöstä olisi prosessille tullut.

Tutkija Neill Boothia kiitän kielentarkastuksen tekemisestä. Ystävällisenä ihmisenä hän otti tehtävän vastaan kiireellisestä aikataulusta huolimatta. Hän paneutui tekstiin huolella ja sain häneltä myös tekstiä selkeyttäviä kommentteja.

Monet ovat olleet ne tahot, jotka ovat rahoittaneet työtäni tutkijana. Kiitokset Tampereen yliopistolle, Syöpäjärjestöille, Kansanterveystieteen tutkijakoululle, Suomen Akatemialle ja Tampereen yliopistolliselle keskussairaalalle.

Ilman suomalaisia miehiä (heidän huolehtivaisia vaimojaan unohtamatta) ei tämä tutkimus olisi tietenkään voinut toteutua. Suomalaiset ovat hyviä osallistumaan tutkimuksiin ja tamperelaiset vielä pääkaupunkiseutulaisia parempia. Kiitän kaikkia miehiä heidän kiinnostuksestaan, ymmärtäväisyydestään ja lojaalisuudestaan näiden vuosien aikana.

Ystäväni, varsinkin jengiläiset, joiden kanssa olen maailmaa parantanut eri yhteyksissä jo 45 vuotta, ovat kiitoksen ansainneet. He ovat jaksaneet uskoa, että tämäkin opiskelujakso elämässäni joskus päätyy positiiviseen tulokseen.

Ensimmäinen Newfoundlandin-koirani Piki ehti olla seuranani tutkimuksen alusta neljä vuotta, ennen kuin antoi periksi korkeassa, lähes 13 vuoden iässä vuonna 2000. Hän tuki kaikessa mielipiteitäni, eikä meille koskaan tullut niistä kiistaa. Hänen seuraajansa Jere, Nöffi hänkin, ei ole niin kiinnostunut sanomisistani, mutta vie minut ulkoilemaan hyvin säännöllisesti pitäen oikeuksistaan kiinni. Lisäksi hän saa minut usein nauramaan ja se piristää raikkaan ilman lisäksi.

Tämän prosessin aikana on perheessäni kaksi sukupolvea saatettu hautaan, kun mummini ja äitini ovat siirtyneet manan majoille, mutta heidän tukensa säilyy hiljaisesti mielessäni. Omat lapseni Elina, Samuli ja Antti ovat tänä aikana kypsyneet oikeiksi aikuisiksi. Perheitä on perustettu ja purettu ja uusi sukupolvi, lapsenlapseni Victor, Frej, Otso ja Ahti, ovat antaneet uutta sisältöä ja vauhtia elämään. Koko tämän, monesti varsin vauhdikkaan elämisenpyörteen keskellä mieheni Taisto on jaksanut kannustaa väitöskirjatyön loppuunsaattamista siinä toivossa, että opiskeluvuodet osaltani vihdoinkin loppuisivat. Eiköhän jo olekin aika mummon päättää opinnot, kun ensimmäinen lapsenlapsi aloittaa koulun.

Helsingissä elokuussa 2007

Liisa Määttänen

9 References

Absetz P, Aro AR and Sutton SR (2003): Experience with breast cancer, pre-screening perceived susceptibility and the psychological impact of screening. Psychooncology 12:305–318.

Ahonen MH, Tenkanen L, Teppo L, Hakama M and Tuohimaa P (2000): Prostate cancer risk and prediagnostic serum 25-hydroxyvitamin D levels (Finland). Cancer Causes Control 11:847–852.

Algarté-Génin M, Cussenot O and Costa P (2004): Prevention of prostate cancer by androgens: experimental paradox or clinical reality. Eur Urol 46:285–294.

Anttila A, Koskela J and Hakama M (2002): Programme sensitivity and effectiveness of mammography service screening in Helsinki, Finland. J Med Screen 9:153–158.

Anttila T, Tenkanen L, Lumme S, Leinonen M, Gislefoss RE, Hallmans G, Thoresen S, Hakulinen T, Luostarinen T, Stattin P, Saikku P, Dillner J, Lehtinen M and Hakama M (2005): Chlamydial antibodies and risk of prostate cancer. Cancer Epidemiol Biomarkers Prev 14:385–389.

Auvinen A, Tammela T, Stenman UH, Uusi-Erkkilä I, Leinonen J, Schröder FH and Hakama M (1996): Screening for prostate cancer using serum prostate-specific antigen: a randomised, population-based pilot study in Finland. Br J Cancer 74:568–572.

Auvinen A, Alexander FE, de Koning HJ and Miller AB (2002a): Should we start population screening for prostate cancer? Randomised trials are still needed. Int J Cancer 97:377–378.

Auvinen A, Määttänen L, Stenman UH, Tammela T, Rannikko S, Aro J, Juusela H and Hakama M (2002b): Lead-time in prostate cancer screening (Finland). Cancer Causes Control 13:279–285.

Auvinen A and Hugosson J (2003): The rationale for the ERSPC trial: will it improve the knowledge base on prostate cancer screening? BJU Int 92 Suppl 2:14–16.

Baade PD, Coory MD and Aitken JF (2004): International trends in prostate-cancer mortality: the decrease is continuing and spreading. Cancer Causes Control 15: 237–241.

Babaian RJ, Mettlin C, Kane R, Murphy GP, Lee F, Drago JR and Chesley A (1992): The relationship of prostate-specific antigen to digital rectal examination and transrectal ultrasonography. Findings of the American Cancer Society National Prostate Cancer Detection Project. Cancer 69:1195–1200.

Boyle P and Ferlay J (2005): Cancer incidence and mortality in Europe, 2004. Ann Oncol 16:481-488.

Brawer MK, Chetner MP, Beatie J, Buchner DM, Vessella RL and Lange PH (1992): Screening for prostatic carcinoma with prostate specific antigen. J Urol 147:841–845.

Brawer MK, Cheli CD, Neaman IE, Goldblatt J, Smith C, Schwartz MK, Bruzek DJ, Morris DL, Sokoll LJ, Chan DW, Yeung KK, Partin AW and Allard WJ (2000): Complexed prostate specific antigen provides significant enhancement of specificity compared with total prostate specific antigen for detecting prostate cancer. J Urol 163:1476–1480.

Brett J, Austoker J and Ong G (1998): Do women who undergo further investigation for breast screening suffer adverse psychological consequences? A multi-centre follow-up study comparing different breast screening result groups five months after their last breast screening appointment. J Public Health Med 20:396–403.

Brooks JD, Metter EJ, Chan DW, Sokoll LJ, Landis P, Nelson WG, Muller D, Andres R and Carter HB (2001): Plasma selenium level before diagnosis and the risk of prostate cancer development. J Urol 166:2034–2038.

Browning DR and Martin RM (2007): Statins and risk of cancer: A systematic review and metaanalysis. Int J Cancer 120:833–843.

Calle EE, Rodriguez C, Walker-Thurmond K and Thun MJ (2003): Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. N Engl J Med 348:1625–1638.

Carter HB, Pearson JD, Metter EJ, Brant LJ, Chan DW, Andres R, Fozard JL and Walsh PC (1992): Longitudinal evaluation of prostate-specific antigen levels in men with and without prostate disease. JAMA 267:2215–2220.

Catalona WJ, Smith DS, Ratliff TL, Dodds KM, Coplen DE, Yuan JJ, Petros JA and Andriole GL (1991): Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. N Engl J Med 324:1156–1161.

Catalona WJ, Smith DS, Ratliff TL and Basler JW (1993): Detection of organ-confined prostate cancer is increased through prostate-specific antigen-based screening. JAMA 270:948–954.

Catalona WJ, Richie JP, Ahmann FR, Hudson MA, Scardino PT, Flanigan RC, deKernion JB, Ratliff TL, Kavoussi LR and 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 151:1283–1290.

Chodak GW, Keller P and Schoenberg HW (1989): Assessment of screening for prostate cancer using the digital rectal examination. J Urol 141:1136–1138.

Chu KC, Tarone RE and Freeman HP (2003): Trends in prostate cancer mortality among black men and white men in the United States. Cancer 97:1507–1516.

Ciatto S, Bonardi R, Mazzotta A, Lombardi C, Santoni R, Cardini S and Zappa M (1995): Comparing two modalities of screening for prostate cancer: digital rectal examination + transrectal ultrasonography vs. prostate-specific antigen. Tumori 81:225–229.

Cole P and Morrison AS (1978): Basic issues in cancer screening. In: Screening in cancer, p 7. Ed AB Miller, UICC Technical Report Series, Vol 40, Geneva.

Day NE and Walter SD (1984): Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. Biometrics 40:1-14.

Day NE (1985): Estimating the sensitivity of a screening test. J Epidemiol Community Health 39:364–366.

Denis L, Murphy GP and Schröder FH (1995): Report of the consensus workshop on screening and global strategy for prostate cancer. Cancer 75:1187–1207.

Dickman PW, Hakulinen T, Luostarinen T, Pukkala E, Sankila R, Soderman B and Teppo L (1999): Survival of cancer patients in Finland 1955-1994. Acta Oncol 12:1–103.

Dorgan JF, Albanes D, Virtamo J, Heinonen OP, Chandler DW, Galmarini M, McShane LM, Barrett MJ, Tangrea J and Taylor PR (1998): Relationships of serum androgens and estrogens to prostate cancer risk: results from a prospective study in Finland. Cancer Epidemiol Biomarkers Prev 7:1069–1074.

Draisma G, Boer R, Otto SJ, van der Cruijsen IW, Damhuis RA, Schröder FH and de Koning HJ (2003): Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst 95:868-–878.

Eichler K, Hempel S, Wilby J, Myers L, Bachmann LM and Kleijnen J (2006): Diagnostic value of systematic biopsy methods in the investigation of prostate cancer: a systematic review. J Urol 175:1605–1612. Review.

Emiliozzi P, Scarpone P, DePaula F, Pizzo M, Federico G, Pansadoro A, Martini M and Pansadoro V (2004): The incidence of prostate cancer in men with prostate specific antigen greater than 4.0 ng/ml: a randomized study of 6 versus 12 core transperineal prostate biopsy. J Urol 171:197–199.

Efstathiou JA, Chen MH, Catalona WJ, McLeod DG, Carroll PR, Moul JW, Roehl KA and D'Amico AV (2006): Prostate-specific antigen-based serial screening may decrease prostate cancer-specific mortality**.** Urology 68:342–327.

Elabbady AA and Khedr MM (2006): Extended 12-core prostate biopsy increases both the detection of prostate cancer and the accuracy of Gleason score. Eur Urol 49:49–53.

Epstein JI, Chan DW, Sokoll LJ, Walsh PC, Cox JL, Rittenhouse H, Wolfert R and Carter HB (1998): Nonpalpable stage T1c prostate cancer: prediction of insignificant disease using free/total prostate specific antigen levels and needle biopsy findings. J Urol 160:2407–2411.

Essink-Bot ML, de Koning HJ, Nijs HG, Kirkels WJ, van der Maas PJ and Schröder FH (1998): Short-term effects of population-based screening for prostate cancer on health-related quality of life. J Natl Cancer Inst 17:925–931.

Etzioni R, Cha R, Feuer EJ and Davidov O (1998): Asymptomatic incidence and duration of prostate cancer. Am J Epidemiol. 148:775–785.

Feinleib M and Zelen M (1969): Some pitfalls in the evaluation of screening programs. Arch Environ Health 19:412–415.

Fink KG, Hutarew G, Pytel A, Esterbauer B, Jungwirth A, Dietze O and Schmeller NT (2003): One 10-core prostate biopsy is superior to two sets of sextant prostate biopsies. BJU Int 92 (Suppl 2):385–388.

Finne P, Zhang WM, Auvinen A, Leinonen J, Maattanen L, Rannikko S, Tammela TL and Stenman UH (2000): Use of the complex between prostate specific antigen and alpha 1-protease inhibitor for screening prostate cancer. J Urol 164:1956–1960.

Freedland SJ and Isaacs WB (2005): Explaining racial differences in prostate cancer in the United States: Sociology or biology? The Prostate 62:243–252.

Fremgen AM, Bland KI, McGinnis LS Jr, Eyre HJ, McDonald CJ, Menck HR and Murphy GP (1999): Clinical highlights from the National Cancer Data Base, CA. Cancer J Clin 49:145–158.

Friedenreich CM, McGregor SE, Courneya KS, Angyalfi SJ and Elliott FG (2004): Case-control study of lifetime total physical activity and prostate cancer risk. Am J Epidemiol 159:740–749.

Gann PH, Hennekens CH and Stampfer MJ (1995): A prospective evaluation of plasma prostate-specific antigen for detection of prostatic cancer. JAMA 273:289–294.

Giles G and Ireland P (1997): Diet, nutrition and prostate cancer. Int J Cancer 10:13–17.

Giles GG, Severi G, English DR, McCredie MR, Borland R, Boyle P and Hopper JL (2003): Sexual factors and prostate cancer. BJU Int 92:211–216.

Giovannucci EL, Liu Y, Leitzmann MF, Stampfer MJ and Willett WC (2005): A prospective study of physical activity and incident and fatal prostate cancer. Arch Intern Med 165:1005–1010.

Gleason DF (1992): Histologic grading of prostate cancer: a perspective. Hum Pathol 23:273–279.

Globocan 2002: Cancer Incidence, Mortality and Prevalence Worldwide IARC CancerBase No.5 version 2.0, IARCPress, Lyon, 2004.

Hakama M (1991): Screening. In: Applications in Public Health, pp. 91–106. Eds. WW Holland, R Detels and G Knox, Oxford Text Book of Public Health, Oxford University Press, Oxford.

Hakama M, Stenman UH, Aromaa A, Leinonen J, Hakulinen T and Knekt P (2001): Validity of the prostate specific antigen test for prostate cancer screening: follow-up study with a bank of 21,000 sera in Finland. J Urol 166:2189–2191.

Hakama M, Auvinen A, Day NE and Miller AB (2007): Sensitivity in screening. Submitted for publication.

Hamdy FC (2007): The prostate cancer prevention trial and its messages. Eur Uro 51:6–8.

Hankey BF, Feuer EJ, Clegg LX, Hayes RB, Legler JM, Prorok PC, Ries LA, Merrill RM and Kaplan RS (1999): Cancer surveillance series: interpreting trends in prostate cancer--part I: Evidence of the effects of screening in recent prostate cancer incidence, mortality, and survival rates. J Natl Cancer Inst 91:1017–1024.

Hoffman RM, Gilliland FD, Adams-Cameron M, Hunt WC and Key CR (2002): Prostate-specific antigen testing accuracy in community practice. BMC Fam Pract 3:19

Holmberg LH, Tabar L, Adami HO and Bergstrom R (1986): Survival in breast cancer diagnosed between mammographic screening examinations. Lancet 2:27–30.

Holund B (1980): Latent prostatic cencer in a consecutive autopsy series. Scand J Urol Nephrol 14:29–43.

Hsing AW, McLaughlin JK, Schuman LM, Bjelke E, Gridley G, Wacholder S, Chien HT and Blot WJ (1990): Diet, tobacco use, and fatal prostate cancer: results from the Lutheran Brotherhood Cohort Study. Cancer Res 50:6836–6840.

Hsing AW, Tsao L and Devesa SS (2000): International trends and patterns of cancer incidence and mortality. Int J Cancer 85:60–67.

Hugosson J, Aus G, Lilja H, Lodding P and Pihl CG (2004): Results of a randomized, population-based study of biennial screening using serum prostate-specific antigen measurement to detect prostate carcinoma. Cancer 100:1397-1405.

Hutchison GB and Shapiro S (1968): Lead time gained by diagnostic screening for breast cancer. J Natl Cancer Inst 41:665–681.

IARC (2002): Breast cancer screening. IARC Handbooks of Cancer Prevention. IARCPress, Lyon.

IARC (2005): Cervix Cancer Screening. IARC Handbooks on Cancer Prevention. IARCPress, Lyon.

Kallio M, Kauraniemi T, Nousiainen AR, Hanstén S, Rytsölä J, Heikkilä M and Hakama M (1994): [Participation of women in cervical cancer screening] [Article in Finnish] Duodecim 110:1061–1067.

Kauppinen A, Kauraniemi T, Koli T and Voipio N (1970): Response to the written invitation in a Gynaecological mass screening by cytology arranged in Helsinki 1969. Acta Obstet Gynecol Scand 7:1–19.

Kerlikowske K, Grady D, Barclay J, Sickles EA and Ernster V (1993): Positive predictive value of screening mammography by age and family history. J Am Med Assoc 275:288–294.

Kilkkinen A, Virtamo J, Virtanen MJ, Adlercreutz H, Albanes D and Pietinen P (2003): Serum enterolactone concentration is not associated with prostate cancer risk in a nested case-control study. Cancer Epidemiol Biomarkers Prev 12:1209–1212.

Klein EA and Thompson IM (2004): Update on chemoprevention of prostate cancer. Curr Opin Urol 14:143–149.

Klein E, Urisman A, Molinaro R, Fischer N, Plummer S, Casey G, Ganem D, DeRisi J, and Silverman R (2006): Identification of a novel retrovirus in prostate tumors of patients homozygous for the R462Q mutation in the HPC1 gene. American Society of Clinical Oncology Symposium. Abstract No: 304

Korodi Z, Dillner J, Jellum E, Lumme S, Hallmans G, Thoresen S, Hakulinen T, Stattin P, Luostarinen T, Lehtinen M and Hakama M (2005): Human papillomavirus 16, 18, and 33 infections and risk of prostate cancer: a Nordic nested case-control study. Cancer Epidemiol Biomarkers Prev 14:2952–2955.

Labrie F, Candas B, Cusan L, Gomez JL, Diamond PS, Suburu R and Lemay M (1996): Diagnosis of noncurable prostate cancer can be practically eliminated by prostate specific antigen. Urology 47:212–217.

Lan SK, Tsai YS, Lin YH and Tzai TS (2007): Diagnostic performance of a random versus lesion-directed biopsy of the prostate from transrectal ultrasound: results of a 5-year consecutive clinical study in 1 institution in South Taiwan. J Ultrasound Med 26:11–17.

Laukkanen JA, Laaksonen DE, Niskanen L, Pukkala E, Hakkarainen A and Salonen JT (2004): Metabolic syndrome and the risk of prostate cancer in Finnish men: a population-based study. Cancer Epidemiol Biomarkers Prev 13:1646–1650.

Lee F, Littrup PJ, Torp-Pedersen ST, Mettlin C, McHugh TA, Gray JM, Kumasaka GH and McLeary RD (1988): Prostate cancer: comparison of transrectal US and digital rectal examination for screening. Radiology 168:389–394.

Legler JM, Feuer EJ, Potosky AL, Merrill RM and Kramer BS (1998): The role of prostate-specific antigen (PSA) testing patterns in the recent prostate cancer incidence decline in the United States. Cancer Causes Control 9:519–527.

Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A and Hemminki K (2000): Environmental and heritable factors in the causation of cancer-analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 343:78–85.

Lilja H, Christenssen A, Dahlen U, Matikainen MT, Nilsson O, Pettersson K and Lövgren T (1991): Prostate-specific antigen in serum occurs predominantly in complex with alpha-1 antichymotrypsin. Clin Chem 37:1618–1625.

MacLennan GT, Eisenberg R, Fleshman RL, Taylor JM, Fu P, Resnick MI and Gupta S (2006): The influence of chronic inflammation in prostatic carcinogenesis: a 5-year follow-up study. J Urol 176:1012–1016.

Malila N, Virtanen MJ, Virtamo J, Albanes D and Pukkala E (2006): Cancer incidence in a cohort of Finnish male smokers. Eur J Cancer Prev 15:103–107.

Matikainen M (2001): Genetic Epidemiology of Hereditary Prostate Cancer in Finland. Doctor's Thesis. University of Tampere. Acta Universitatis Tamperensis No. 812. Tampere.

Mettlin C, Murphy GP, Babaian RJ, Chesley A, Kane RA, Littrup PJ, Mostofi FK, Ray PS, Shanberg AM and Toi A (1996): The results of a five-year early prostate cancer detection intervention. Investigators of the American Cancer Society National Prostate Cancer Detection Project. Cancer 77:150–159.

Miller AB, Chamberlain J, Day NE, Hakama M and Prorok PC (eds.) (1991): Cancer screening. UICC Project on Evaluation of Screening for Cancer. International Union Against Cancer, Cambridge.

Miller BA, Kolonel LN and Bernstein L, et al., eds. (1996): Racial/Ethnic Patterns of Cancer in the United States 1988–1992. NIH Publication No. 96–4104. National Cancer Institute, Bethesda, MD.

Mistry K and Cable G (2003): Meta-analysis of prostate-specific antigen and digital rectal examination as screening tests for prostate carcinoma. J Am Board Fam Pract 16:95–101.

Mäkinen T, Tammela TL, Hakama M, Stenman UH, Rannikko S, Aro J, Juusela H, Määttänen L and Auvinen (2001): 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 166:1339–1342.

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

Mäkinen T, Tammela TL, Stenman UH, Määttänen L, Rannikko S, Aro J, Juusela H, Hakama M and Auvinen A (2002b): Family history and prostate cancer screening with prostate-specific antigen. J Clin Oncol 20:2658-2663.

Nelen V, Thys G, Neels H, Neetens I, Coebergh JW, Dourcy-Belle-Rose B and Denis L (2003): ERSPC: features and preliminary results from the Antwerp study centre. BJU Int 92 Suppl 2:17-21.

Nelson WG, De Marzo AM and Isaacs WB (2003): Prostate cancer. N Engl J Med 349:366–381.

Norberg M, Egevad L, Holmberg L, Sparen P, Norlen BJ and Busch C (1997): The sextant protocol for ultrasound-guided core biopsies of the prostate underestimates the presence of cancer. Urology 50:562–566.

Oesterling JE, Jacobsen SJ, Klee GG, Pettersson K, Piironen T, Abrahamsson PA, Stenman UH, Dowell B, Lövgren T and Lilja H (1995): Free, complexed and total serum prostate specific antigen: the establishment of appropriate reference ranges for their concentrations and ratios. J Urol 154:1090–1095.

Oliver SE, May MT and Gunnell D (2001): International trends in prostate-cancer mortality in the 'PSA era'. Int J Cancer 92:893–898.

Parkes C, Wald NJ, Murphy P, George L, Watt HC, Kirby R, Knekt P, Helzlsouer KJ and Tuomilehto J (1995): Prospective observational study to assess value of prostate specific antigen as screening test for prostate cancer. BMJ 311:1340–1343.

Parkin DM, Whelan SL, Ferlay J, Teppo L and Thomas DB (eds) (2002): Cancer Incidence in Five Continents, International Agency for Research on Cancer (IARC) Scientific Publications No. 155, IARC Press, Lyon.

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

Parsons JK, Carter HB, Platz EA, Wright EJ, Landis P and Metter EJ (2005): Serum testosterone and the risk of prostate cancer: potential implications for testosterone therapy. Cancer Epidemiol Biomarkers Prev 14:2257–2260.

Plaskon LA, Penson DF, Vaughan TL and Stanford JL (2003): Cigarette smoking and risk of prostate cancer in middle-aged men. Cancer Epidemiol Biomarkers Prev 12:604–609.

Potosky AL, Kessler L, Gridley G, Brown CC and Horm JW (1990): Rise in prostatic cancer incidence associated with increased use of transurethral resection. J Natl Caner Inst 82:1624–1628.

Potosky AL, Miller BA, Albertsen PC and Kramer BS (1995): The role of increasing detection in the rising incidence of prostate cancer. JAMA 273:548–552

Richie JP, Catalona WJ, Ahmann FR, Hudson MA, Scardino PT, Flanigan RC, deKernion JB, Ratliff TL, Kavoussi LR, Dalkin BL, et al. (1993): Effect of patient age on early detection of prostate cancer with serum prostate-specific antigen and digital rectal examination. Urology 42:365–374.

Ries LAG, Harkins D, Krapcho M, Mariotto A, Miller BA, Feuer EJ, Clegg L, Eisner MP, Horner MJ, Howlader N, Hayat M, Hankey BF and Edwards BK (eds) (2006): SEER Cancer Statistics Review, 1975-2003, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2003/, based on November 2005 SEER data submission, posted to the SEER web site 2006.

Rietbergen JB, Kranse R, Hoedemaeker RF, Kruger AE, Bangma CH, Kirkels WJ and Schröder FH (1998): Comparison of prostate-specific antigen corrected for total prostate volume and transition zone volume in a population-based screening study. Urology 52:237–246.

Rintala P, Pukkala E, Laara E and Vihko V (2003): Physical activity and breast cancer risk among female physical education and language teachers: a 34-year follow-up. Int J Cancer107:268–270.

Ritch CR, Wan RL, Stephens LB, Taxy JB, Huo D, Gong EM, Zagaja GP and Brendler CB (2007): Dietary fatty acids correlate with prostate cancer biopsy grade and volume in Jamaican men. J Urol 177:97–101.

Roobol MJ, Zappa M, Määttänen L and Ciatto S (2007): The value of different screening tests in predicting prostate biopsy outcome in screening for prostate cancer data from a multicenter study (ERSPC). Prostate 67:439–446.

Ross R, Bernstein L, Judd H, Hanisch R, Pike M and Henderson B (1986): Serum testosterone levels in healthy young black and white men. J Natl Cancer Inst 76:45–48.

Sakr WA, Haas GP, Cassin BF, Pontes JE and Crissman JD (1993): The frequency of carcinoma and intraepithelial neoplasia of the prostate in young male patients. J Urol 150:379–385.

Sakr WA, Grignon DJ, Crissman JD, Heilbrun LK, Cassin BJ, Pontes JJ and Haas GP (1994): High grade prostatic intraepithekial neoplasia (HGPIN) and prostatic adenocarcinoma between the ages of 20-69: an autopsy study of 249 cases. In Vivi 8:439–443.

Samanta M, Harkins L, Klemm K, Britt WJ and Cobbs CS (2003): High prevalence of human cytomegalovirus in prostatic intraepithelial neoplasia and prostatic carcinoma. J Urol 170:998–1002.

Schröder FH(1993): Prostate cancer: to screen or not to screen? BMJ 306:407–408.

Schröder FH, Damhuis RAM, Kirkels WJ, De Koning HJ, Kranse R, Nisj HGT and Blijenberg BG (1996): European randomized study of screening for prostate cancer-the Rotterdam pilot studies. Int J Cancer 65:145-151.

Schröder FH, van der Maas P, Beemsterboer P, Kruger AB, Hoedemaeker R, Rietbergen J and 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 90:1817–1823.

Schröder FH, Denis LJ, Roobol M, Nelen V, Auvinen A, Tammela T, Villers A, Rebillard X, Ciatto S, Zappa M, Berenguer A, Paez A, Hugosson J, Lodding P, Recker F, Kwiatkowski M and Kirkels WJ; ERSPC (2003): The story of the European Randomized Study of Screening for Prostate Cancer. BJU Int 92 (Suppl 2):1–13.

Singh H, Canto EI, Shariat SF, Kadmon D, Miles BJ, Wheeler TM and Slawin KM (2004): Improved detection of clinically significant, curable prostate cancer with systematic 12-core biopsy. J Urol 171:1089–1092.

Siu W, Dunn RL, Shah RB and Wei JT (2005): Use of extended pattern technique for initial prostate biopsy. J Urol 174:505–509.

Smith DS and Catalona WJ (1994): Rate of change in serum prostate specific antigen levels as a method for prostate cancer detection. J Urol 152:1163–1167.

Smith DC, Catalona WJ and Herschman JD (1996): Longitudinal screening for prostate cancer with prostate specific antigen. J Am Med Assoc 276:1309–1315.

Stattin P, Adlercreutz H, Tenkanen L, Jellum E, Lumme S, Hallmans G, Harvei S, Teppo L, Stumpf K, Luostarinen T, Lehtinen M, Dillner J and Hakama M (2002): Circulating enterolactone and prostate cancer risk: a Nordic nested case-control study. Int J Cancer 99:124–129.

Stattin P, Lumme S, Tenkanen L, Alfthan H, Jellum E, Hallmans G, Thoresen S, Hakulinen T, Luostarinen T, Lehtinen M, Dillner J, Stenman UH and Hakama M (2004): High levels of circulating testosterone are not associated with increased prostate cancer risk: a pooled prospective study. Int J Cancer 108:418–424.

Stenman U-H, Leinonen J, Alfthan H, Rannikko S, Tuhkanen K and Alfthan O (1991): 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 51:222–226.

Stenman UH, Hakama M, Knekt P, Aromaa A, Teppo L and Leinonen J (1994): Serum concentrations of prostate specific antigen and its complex with alpha 1 –antichymotrypsin before diagnosis of prostate cancer. Lancet 344:1594–1598.

Sunny L (2005): Prostate Cancer – An Epidemiological Study in India. Doctor's Thesis. University of Tampere. Acta Universitatis Tamperensis No. 1098. Tampere.

Thompson IM, Rounder JB, Teague JL, Peek M and Spence CR (1987): Impact of routine screening for adenocarcinoma of the prostate on stage distribution. J Urol. 137:424-426.

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 and Coltman CA Jr (2003): The influence of finasteride on the development of prostate cancer. N Engl J Med 349:215–224.

Thompson IM, Ankerst DP, Chi C, Goodman PJ, Tangen CM, Lucia MS, Feng Z, Parnes HL and Coltman CA Jr. (2006): Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial. J Natl Cancer Inst 98:529–534.

Tollefson MK, Leibovich BC, Slezak JM, Zincke H and Blute ML (2006): Long-term prognostic significance of primary Gleason pattern in patients with Gleason score 7 prostate cancer: impact on prostate cancer specific survival. J Urol 175:547–551.

Tuohimaa P, Tenkanen L, Ahonen M, Lumme S, Jellum E, Hallmans G, Stattin P, Harvei S, Hakulinen T, Luostarinen T, Dillner J, Lehtinen M and Hakama M (2004a): Both high and low levels of blood vitamin D are associated with a higher prostate cancer risk: a longitudinal, nested case-control study in the Nordic countries. Int J Cancer 108:104–108.

Tuohimaa P, Tenkanen L, Ahonen M, Lumme S, Jellum E, Hallmans G, Stattin P, Harvei S, Hakulinen T, Luostarinen T, Dillner J, Lehtinen M and Hakama M (2004b): Metabolism of 25-hydroxyvitamin D3 may explain the u-shaped risk curve for prostate cancer. Int J Cancer 111:469.

Tuohimaa P, Tenkanen L, Syvälä H, Lumme S, Hakulinen T, Dillner J and Hakama M (2007): Interaction of factors related to the metabolic syndrome and vitamin D on risk of prostate cancer. Cancer Epidemiol Biomarkers Prev 16:302–307.

United Kingdom Trial of Early Detection of Breast Cancer Group (1992): Specificity of screening in United Kingdom trial of early detection of breast cancer. Br Med J 304:346–349.

Virtamo J, Pietinen P, Huttunen JK, Korhonen P, Malila N, Virtanen MJ, Albanes D, Taylor PR, Albert P and ATBC Study Group (2003): Incidence of cancer and mortality following alpha-tocopherol and beta-carotene supplementation: a postintervention follow-up. JAMA 290:476–485.

Vo T, Rifkin MD and Peters TL (2001): Should ultrasound criteria of the prostate be redefined to better evaluate when and where to biopsy. Ultrasound Q 17:171–176.

Walter SD (1999): Estimation of test sensitivity and specificity when disease confirmation is limited to positive results. Epidemiology 10:67–72.

Wang MC, Valenzuela LA, Murphy GP and Chu TM (1979): Purification of a human prostate specific antigen. Invest Urol 17:159–163.

Whittemore AS, Kolonel LN, Wu AH, John EM, Gallagher RP, Howe GR, et al. (1995): Prostate cancer in relation to diet, physical activity, and body size in blacks, whites, and Asians in the United States and Canada. J Natl Cancer Inst 87:652–661.

Wilson JMG and Jungner G (1968): Principles and practice of screening for disease. Public health paper no. 34. World Health Organization, Geneva.

www.cancerregistry.fi/eng/, Basic statistics, Finnish Cancer Registry, 4.4.2007.

www.seer.cancer.gov, Surveillance, epidemiology, and end results (SEER) program. 2006.

Zeegers MP, Dirx MJ and van den Brandt PA (2005): Physical activity and the risk of prostate cancer in the Netherlands cohort study, results after 9.3 years of follow-up. Cancer Epidemiol Biomarkers Prev 14:1490–1495.

Original Publications

BRIEF COMMUNICATION

Three-Year Results of the Finnish Prostate Cancer Screening Trial

Liisa Määttänen, Anssi Auvinen, Ulf-Håkan Stenman, Teuvo Tammela, Sakari Rannikko, Jussi Aro, Harri Juusela, Matti Hakama

Prostate cancer screening is increasing despite lack of demonstrated effectiveness *(1)*. Large randomized, controlled trials with long-term follow-up are ongoing in Europe and North America to assess the effects of screening on mortality and quality of life *(2)*.

The Finnish trial is a part of the European Randomized Study of Screening for Prostate Cancer, a multicenter trial with eight participating centers *(3)*. The common core protocol includes enrollment of men at age 55–67 years and a screening test to assay the concentration of prostate-specific antigen (PSA) in serum, with a PSA cutoff point of 4.0 ng/mL.

We report here the attendance rate, the specificity, and the detection rate of prostate cancer during the first 3 years of the prevalence screening round. These parameters are intermediate indicators and thus are necessary, but are not sufficient, conditions for effective screening.

The target population of the Finnish prostate cancer screening trial consists of men born during the period from 1929 through 1944 who reside in the metropolitan areas of Helsinki or Tampere, Finland. During the first 3 years of the study (1996–1998), 60 211 men aged 55–67 years were identified from the Population Registry of Finland. Information on prostate cancers was obtained through a record linkage with the Finnish Cancer Registry, and men with prevalent prostate cancer were excluded from the study before randomization (n = 238).

Annually, 8000 men were randomly assigned to the screening arm, and the roughly 12 000 men remaining in the target population (a total of 35 973) were randomly assigned to the control arm. Men in the screening arm were recruited by invitations that were mailed in four batches annually. Men who were deceased, had moved outside the study area, or had prohibited the use of their addresses between randomization and the date of mailing were considered to be ineligible (n = 1268). After written informed consent was obtained from the men, a blood sample was drawn from the men in the screening group. The concentration of PSA in serum was determined with the Tandem-E assay (Hybritech, San Diego, CA) or, in case of equipment malfunction, with another assay calibrated to the Tandem-E assay.

Men with a serum PSA concentration of 4.0 ng/mL or higher were referred to diagnostic examinations. These examinations consisted of digital rectal examination (DRE), transrectal ultrasound, and transrectal prostate biopsy examination.

Men with a PSA concentration of 3.0–3.9 ng/mL were offered a DRE by a urologist. Initially, 119 men with a PSA concentration of 2.0–2.9 ng/mL were also offered a DRE, but this was soon discontinued because of its poor efficiency and high cost. Men with a suspicious DRE finding, however, were referred to the other diagnostic examinations described above.

The positive predictive value (PPV) was estimated as the number of cancers detected among men with defined PSA concentrations, divided by the number of men within the PSA range. The detection rate was calculated as the prevalence of screening-detected cancers, i.e., the number of cancers among men with given criteria, relative to the number of men fulfilling those criteria. Specificity was estimated as the proportion of men with negative screening test results among men without prostate cancer.

The study protocol was approved by an ethical committee in each participating hospital.

Of the 22 732 eligible men in the screening arm, 69% (15 685 men) participated (Table 1). The participation rate did not vary substantially by age. The serum PSA concentration was 3.0 ng/mL or greater in 14% (2143 men) of the participants in the screening arm. As expected, the PSA concentration increased with age. At baseline, 10% (1334 of 13 802 men, 1883 with missing information) reported prior PSA screening and 5% (719 of 13 240 men, 2445 with missing information) reported a first-degree relative affected with prostate cancer.

The overall detection rate of the screening program was 2.6% (Table 2). The detection rates ranged from 1% at 55 years of age to 5% at 67 years of age. The detection rate was 2.1% for cancers of Gleason grades 2–6 and 0.4% for cancers of Gleason grades 7–10. (The Gleason grade was unavailable for seven patients with prostate cancer.)

The specificity of the PSA test increased rapidly with increasing cutoff level. Of the 1342 men with a PSA concentration of 4.0 ng/mL or greater, 1236 (92%) underwent diagnostic examinations. A total of 386 men with a PSA concentration of 4.0 ng/mL or greater were diagnosed with prostate cancer. The detection rate attributable to a PSA concentration of 4.0 ng/mL or greater was 2.5%, and the specificity was 93%. A PSA cutoff point of 10 ng/mL gave a higher specificity (99%), but the detection rate was only 0.9%. The PPV was 22% for the PSA range of 4.0–9.9 ng/mL and 62% for a PSA concentration of 10 ng/mL or higher. A DRE was offered to 801 men with a PSA concentration of 3.0–3.9 ng/mL, 92% (733) of them complied, and 22 cancers were detected. This practice contributed modestly to the detection rate (0.1%) and lowered specificity (from 93% to 88%).

The Finnish prostate cancer screening trial is population based. Hence, the results are generalizable to screening as a public health policy, unlike results

Affiliations of authors: L. Määttänen, Finnish Cancer Registry, Helsinki, Finland; A. Auvinen, Tampere School of Public Health, University of Tampere, Finland; U.-H. Stenman, Department of Clinical Chemistry, Helsinki University Hospital; T. Tammela, Department of Urology, Tampere University Hospital; S. Rannikko, Department of Urology, Helsinki University Hospital; J. Aro, Department of Surgery, Helsinki City Hospital; H. Juusela, Department of Surgery, Jorvi Hospital, Espoo, Finland; M. Hakama, Finnish Cancer Registry and Tampere School of Public Health, University of Tampere.

Correspondence to: Liisa Määttänen, M.P.H., Finnish Cancer Registry, Liisankatu 21 B, 00170 Helsinki, Finland (e-mail: liisa.maattanen@cancer.fi).

See "Notes" following "References."

Table 1. Number of eligible men, participation, and serum prostate-specific antigen (PSA) concentration by age in the Finnish prostate cancer screening trial, 1996–1998

Age, y	Eligible, No.	Participated, No.	PSA concentration, ng/mL <2 No.	<2 %*	2.0–2.9 No.	2.0–2.9 %*	3.0–3.9 No.	3.0–3.9 %*	4.0–9.9 No.	4.0–9.9 %*	⩾10 No.	⩾10 %*
55	7308	4833	4113	85	381	8	159	3	159	3	21	0
59	6041	4238	3364	79	397	9	186	4	244	6	47	1
63	4901	3503	2473	71	416	12	221	6	324	9	69	2
67	4482	3111	1958	63	440	14	235	8	389	13	89	3
Total	22 732	15 685	11 908	76	1634	10	801	5	1116	7	226	1

*Percentage of participants.

Table 2. Cumulative number of men and prostate cancers, specificity (with 95% confidence interval [CI]) and detection rate (with 95% CI) by serum concentration of prostate-specific antigen (PSA) in the Finnish prostate cancer screening trial, 1996–1998

PSA concentration, ng/mL	No. of men	No. of cancers	Positive predictive value, % (95% CI)	Specificity,* % (95% CI)	Detection rate,* % (95% CI)
<2	11 908	—	—	—	2.6 (2.4 to 2.9)
2.0–2.9†	1634	3	3 (1 to 8)	78 (77 to 79)	2.6 (2.4 to 2.9)
3.0–3.9†	801	22	3 (2 to 4)	88 (88 to 89)	2.6 (2.4 to 2.9)
4.0–9.9	1116	247	22 (20 to 25)	93 (93 to 94)	2.5 (2.2 to 2.7)
⩾10	226	139	62 (55 to 68)	99 (99 to 99)	0.9 (0.7 to 1.0)
Total	15 685	411	NA‡	100	2.6 (2.4 to 2.9)

*For PSA concentrations above the lower limit of the class.

†Combination of digital rectal examination and PSA assay as the screening test for PSA level of 3.0–3.9 ng/mL; in addition, 108 men with a PSA concentration of 2.0–2.9 ng/mL underwent a digital rectal examination in 1996.

‡NA = not applicable.

from volunteer-based efficacy trials. The high participation rate (69%) indicates that, in Finland, prostate cancer screening will be feasible, if it is found to be effective. At baseline, opportunistic screening was relatively low in Finland and does not seem to jeopardize the trial.

Specificity determines the costs and acceptability of a prostate cancer screening program. To increase specificity, we used a screening algorithm with a relatively high PSA cutoff point and few auxiliary interventions. Specificity and PPV increased rapidly with increasing concentrations of serum PSA, indicating less harm and lower costs per screen-detected cancer at higher cutoff levels. We found a PPV for serum PSA concentrations of 4.0 ng/mL or greater of 29%, i.e., 3.5 biopsy examinations per cancer. Lowering the PSA cutoff point to 3 ng/mL and abandoning the DRE would have increased referral to diagnostic examinations by 5%, in addition to the 9% of men with a PSA concentration of 4 ng/mL or greater.

The detection rate for cancers of Gleason grades 7–10 was 0.4%, which indicates that the screening program also is able to detect clinically significant cancers. However, information on interval cancer incidence and the detection rate at a second screening round are required for assessment of overdiagnosis and sensitivity.

In summary, the Finnish prostate cancer screening trial demonstrates that screening is acceptable for the target population, the performance of the screening test is adequate, and the detection rate of aggressive, potentially lethal cancer is reasonable. These results pertaining to intermediate indicators provide necessary, but not sufficient, indication for the effectiveness of prostate cancer screening.

REFERENCES

(1) The International Prostate Screening Trial Evaluation Group. Rationale for randomised trials of prostate cancer screening. Eur J Cancer 1999;35:262–71.

(2) Auvinen A, Rietbergen JB, Denis LJ, Schroder FH, Prorok PC. Prospective evaluation plan for randomised trials of prostate cancer screening. The International Prostate Cancer Screening Trial Evaluation Group. J Med Screen 1996;3:97–104.

(3) Maattanen L, Auvinen A, Stenman UH, Rannikko S, Tammela T, Aro J, et al. European randomized study of prostate cancer screening: first-year results of the Finnish trial. Br J Cancer 1999;79:1210–4.

NOTES

Supported by the Academy of Finland, the Cancer Society of Finland, the Helsingin Sanomat Centenarian Foundation, the Tampere University Hospital Research Fund, the Finnish Cultural Fund, the Hybritech Corporation, and the Europe Against Cancer Program.

We are grateful to the Pirkanmaa Cancer Society (Tampere, Finland) and the Aleksi Medical Center (Helsinki, Finland) for drawing the blood samples for the study, to Mr. Jarkko Koivuniemi for programming, and to Ms. Minna Heikkilä for data management.

Manuscript received September 18, 2000; revised November 8, 2000; accepted January 23, 2001.

British Journal of Cancer (2007) 96, 56–60

www.bjcancer.com

Specificity of serum prostate-specific antigen determination in the Finnish prostate cancer screening trial

L Määttänen[1,2], M Hakama[1,2], TLJ Tammela[3], M Ruutu[4], M Ala-Opas[4], H Juusela[5], P Martikainen[6], U-H Stenman[7] and A Auvinen[*,2,8]

[1]Finnish Cancer Registry, Liisankatu 21 B, FIN-00170 Helsinki, Finland; [2]Tampere School of Public Health, FIN-33014 University of Tampere, Tampere, Finland; [3]Department of Urology, Tampere University Hospital and University of Tampere, Box 2000, FIN-33521 Tampere, Finland; [4]Department of Urology, Helsinki University Hospital, Box 580 FIN-00029, Helsinki, Finland; [5]Department of Surgery, Jorvi Hospital, Turuntie 150, FIN-02740 Espoo, Finland; [6]Department of Pathology, Tampere University Hospital, University of Tampere, Box 2000, FIN-33521 Tampere, Finland; [7]Department of Clinical Chemistry, Helsinki University Hospital, Box 700, FIN-00029 Helsinki, Finland; [8]Finnish Cancer Institute, Liisankatu 21 B, FIN-00170 Helsinki, Finland

Specificity constitutes a component of validity for a screening test. The number of false-positive (FP) results has been regarded as one of major shortcomings in prostate cancer screening. We estimated the specificity of serum prostate-specific antigen (PSA) determination in prostate cancer screening using data from a randomised, controlled screening trial conducted in Finland with 32 000 men in the screening arm. We calculated the specificity as the proportion of men with negative findings (screen negatives, SN) relative to those with negative and FP results (SN/(SN + FP)). A SN finding was defined as either PSA $\leqslant 4\,ng\,ml^{-1}$ or PSA 3.0–3.9 $ng\,ml^{-1}$ combined with a negative ancillary test (digital rectal examination, DRE or free/total, F/T PSA ratio). False positives were those with positive screening test followed by a negative diagnostic examination. Of the 30 194 eligible men, 20 794 (69%) attended the first screening round and 1968 (9.5%) had a screen-positive finding. A total of 508 prostate cancers were detected at screening (2.4%). Hence, the number of SN findings was 18 825 and the number of FP results 1358. Specificity was estimated as 0.933 (18 825 out of 20 183) with 95% confidence interval (CI) 0.929–0.936. Specificity decreased with age. Digital rectal examination as ancillary examination had similar or higher specificity than F/T PSA. In the second screening round, specificity was slightly lower (0.912, 95% CI 0.908–0.916). The specificity of PSA screening in the Finnish screening trial is acceptable. Further improvement in specificity could, however, improve acceptability of screening and decrease screening costs.
British Journal of Cancer (2007) **96,** 56–60. doi:10.1038/sj.bjc.6603522 www.bjcancer.com

Keywords: prostate neoplasms; mass screening; specificity; randomised controlled trial; prostate-specific antigen

The validity of a screening test is the capability to discriminate between those with and without disease, and can be measured by sensitivity and specificity (Hakama, 1991). Ideally, a screening test should be able to classify correctly both subjects with and without the target disorder. In practice, the distributions of test values between these two populations always overlap. Sensitivity indicates the capacity to find persons with disease, whereas specificity is the ability to identify those free of the target disorder. Sensitivity and specificity are characteristics of the test that are independent of the occurrence of the disease in the target population, but may depend on disease characteristics. Specificity is inversely proportional to the frequency of FP tests in those free of disease. In the context of cancer screening, optimal specificity depends on how many negative (unnecessary) biopsies one is willing to accept in order to detect one case of cancer. Specificity of a screening test is an indicator of the adverse effects of screening, including the cost and inconvenience owing to the diagnostic examination. Specificity is a characteristic of the test and an indicator of test performance used in evaluation of screening methods, but not directly applicable in decision making at the individual level.

Prostate cancer is one of the most common cancers among men in the industrial countries (Parkin *et al*, 2002). Serum prostate-specific antigen (PSA) was identified in the 1970s (Ablin *et al*, 1970; Li and Beling 1973) and later shown to be a marker of prostate cancer (Wang *et al*, 1981). It has been adopted for case finding among asymptomatic men, which has substantially increased the detection and incidence of prostate cancer (Hankey *et al*, 1999; Etzioni *et al*, 2002). Opportunistic screening with PSA is widespread, but the evidence for its effectiveness in terms of mortality reduction is still lacking (Auvinen *et al*, 2002). One of the problems with PSA screening is the large proportion of FP results, as PSA is an organ-specific, but not disease-specific marker (Stenman *et al*, 2000). The main cause of elevated serum PSA concentration is benign prostatic hyperplasia. A positive screening test in the absence of disease leads to unnecessary biopsies and constitutes an adverse effect of screening, which adds costs, increases overdiagnosis and overtreatment and can affect acceptability of screening, that is, reduce participation at subsequent

*Correspondence: Professor A Auvinen, Finnish Cancer Institute, Liisankatu 21 B, FIN-00170, Helsinki, Finland;
E-mail: anssi.auvinen@uta.fi
Received 8 September 2006; revised 10 November 2006; accepted 13 November 2006

screening rounds. This is especially important in population screening, where the proportion of those with disease is low.

The aim of the study was to estimate the specificity of the PSA test in the Finnish prostate cancer screening trial.

MATERIALS AND METHODS

The Finnish prostate cancer screening trial is the largest centre of the European Randomised Study of Prostate Cancer screening (de Koning *et al*, 2002). It was started in 1996 in two metropolitan regions, Helsinki and Tampere. The study population of 80 458 men at ages 55–67 years was identified from the Population Register Centre of Finland. Men who had denied the use of their addresses were ineligible (approximately 1%), as well as men with a previous prostate cancer ($N = 161$) and they were excluded from the trial. During 1996–1999, 8000 men were annually randomly allocated to the screening arm using a computer algorithm based on random numbers and were invited for the first screening round. The second screening round was carried out after a 4-year interval between 2000 and 2003. The rest of the target population comprised the control arm of the trial. Individuals in the control arm were not contacted.

An invitation letter was sent to the men of the screening arm with an information leaflet describing the trial, appended with a brief questionnaire about urological symptoms, family history of prostate cancer, previous PSA tests and an informed consent form to be signed by the subject. This approach is called randomisation before consent or Zelen-type randomisation (Zelen, 1979).

After an informed consent, a blood sample was drawn at the local cancer society clinics in Helsinki and Tampere. Serum PSA concentrations were analysed at the Central Laboratory of Helsinki University Hospital by Hybritech Tandem-E for determination of total PSA and Wallac AutoDelfia for free PSA.

Men with serum PSA $\geqslant 4\,\mathrm{ng\,ml^{-1}}$ were referred to the local hospital for diagnostic examinations, including three examinations for all men: digital rectal examination (DRE), transrectal ultrasound (TRUS) and prostate biopsy (compliance with biopsy 95%). Men with serum PSA concentration 3.0–3.9 $\mathrm{ng\,ml^{-1}}$ were referred for supplementary test: DRE during the first 3 years and the proportion of free PSA (F/T-PSA) since 1999 with a cutoff point of 0.16. Those with a positive ancillary test were also referred to diagnostic work-up. Diagnosis of prostate cancer was based on histological confirmation. The biopsy protocol consisted initially of sextant biopsies, but the number of cores was increased to 10–12 in 2002. A re-biopsy was carried out if either the PSA was above 10 $\mathrm{ng\,ml^{-1}}$ or the initial histopathologic diagnosis was prostatic intraepithelial neoplasia, atypical small acinar proliferation or unconfirmed suspicion for carcinoma. Information on screen-detected cases was obtained from the trial database and interval cancers were identified from the population-based, nationwide Finnish Cancer Registry (Teppo *et al*, 1994).

Specificity was defined as the proportion of the disease-free men correctly classified as negative by the test (PSA $< 4.0\,\mathrm{mg\,ml^{-1}}$, or PSA 3.0–3.9 $\mathrm{ng\,ml^{-1}}$ with a positive ancillary test), among men classified as disease-free during the screening episode (including both SN men and those who were screen positive but had a negative biopsy). Confidence interval (CI) for specificity was calculated based on standard error for a proportion, s.e. $(p) = \sqrt{(pq/n)}$.

The study protocol was approved by the ethical committees in each participating hospital. Permission to use the data of the cancer registry was obtained from the Research and Development Center for Welfare and Health (STAKES).

RESULTS

In the target population of 80 458 men, 8000 men were randomly allocated to the screening arm each year during the enrolment period, 1996–1999. At the time of the invitation to the first screening round, 30 194 were eligible and invited. Of them, 20 794 (69%) participated in the first screening round. A drop-out analysis showed that young age and residence in Helsinki area were associated with non-participation (mean ages 59.8 *vs* 60.2 years and 81% *vs* 72% resident in Helsinki region among non-participants and participants). Of the participants, 18,825 had serum PSA concentration below 3 $\mathrm{ng\,ml^{-1}}$ or PSA 3.0–3.9 $\mathrm{ng\,ml^{-1}}$ in combination with negative DRE or free/total ratio $\geqslant 0.16$ (Table 1). Thus, the number of screen-positive men was 1968 and prostate cancer was histologically confirmed in 508 subjects. The histological finding was benign in 1358 men and 102 subjects (5.2%) did not undergo biopsy.

Based on these observations, the specificity of the screening test was estimated as $(18\,825)/(18\,825 + 1358)$ that is, 0.933 with 95% CI 0.929–0.936 (Table 2). Men without biopsy were excluded from this calculation. Specificity decreased with increasing age, from 0.97 at age 55 to 0.88 at 67 years.

Assuming a similar proportion of cancers and benign findings among the 102 men who were not biopsied as among the screen-positive men who underwent biopsy $(102 \bullet (508/\{508 + 1{,}358\}))$, the number of cancers was estimated as 28 and number of men with FP screening test as 74. Therefore, a corrected estimate of the relative specificity of the PSA test was virtually identical to the original: $(18\,825)/(18\,825 + 1432) = 0.929$ (95% CI 0.926–0.933).

After the first 3 years of screening, DRE was replaced with F/T PSA ratio as ancillary test among men with PSA 3.0–3.9 $\mathrm{ng\,ml^{-1}}$. The proportion of FP results among men with PSA in this range was 7% with DRE (59 out of 794) and 16% with F/T PSA ratio > 0.16 (44 out of 269). The overall number of SN findings during the three initial years was 14 149 and the number of FP findings 995 (7.0%). During the last year of the first screening round, the corresponding figures were 4540 and 368 (8.1%). Hence, adoption of the F/T PSA to replace DRE was associated with a nonsignificant

Table 1 Number of men by screening result and prostate cancer diagnosis in the first screening round, Finnish prostate cancer screening trial

	Prostate cancer diagnosis		
Screening result	**Yes**	**No**	**Total**
Positive	508[a]	1358	1866[b]
Negative	42[c]	18 783	18 825
Total	550	20 141	20 691[b]

[a]Screen-detected prostate cancer. [b]A total of 102 screen-positive men failed to undergo biopsy and were excluded. [c]Interval cancer among screen-negative men.

Table 2 The frequency of TN and FP screening findings by age in the first screening round of the Finnish prostate cancer screening trial

Age	TN[a]	FP[b]	Specificity (95% CI)[c]
55	6153	214	0.966 (0.962–0.971)
59	5137	279	0.948 (0.943–0.954)
63	4146	409	0.910 (0.902–0.919)
67	3389	456	0.881 (0.871–0.892)
Total	18 825	1358	0.933 (0.929–0.936)

Abbreviations: CI = confidence interval; FP = false positive; TN = true negative. [a]TN: No. of men with negative screening result (serum PSA $< 3.0\,\mathrm{ng\,ml^{-1}}$ or PSA 3.0–3.9 $\mathrm{ng\,ml^{-1}}$ with a negative ancillary examination (benign finding at digital rectal examination or free/total PSA ratio $\geqslant 0.16$)). [b]FP: No. of men with positive screening result (serum PSA $< 4.0\,\mathrm{ng\,ml^{-1}}$ or PSA 3.0–3.9 $\mathrm{ng\,ml^{-1}}$ with a positive ancillary examination (suspicious finding at digital rectal examination or free/total PSA ratio < 0.16)) minus number of screen-detected cancers. Note: men refusing biopsy ($N = 102$) excluded. [c]Specificity: TN/(TN+FP).

decrease in specificity of the screening programme, 0.934 (0.930–0.938) *vs* 0.925 (0.917–0.932) in the first screening round.

Specificity in the second round was slightly lower compared with the first round (Table 3). A total of 18 612 men were screened and 2303 screen-positive subjects referred to biopsy. Of them, 2156 were biopsied within the study (at the screening centres) and 583 cancers (3.1%) detected. Overall, specificity was 0.912 (95% CI 0.908–0.916, Table 4). Correction for missing biopsy results did not materially affect the estimate (corrected specificity 0.910, 95% CI 0.906–0.914). In men who attended screening for the first time (i.e., were non-participants in the first round), specificity was 0.903 (95% CI 0.891–0.914), with a corrected estimate of 0.911 (0.900–0.921). Similar to the first screening round, specificity decreased with age. However, no obvious difference in specificity was found within age group, that is, when comparing men at the same age in the first *vs* second round. Specificity in the two screening rounds remained comparable (0.917 and 0.922) after restricting the analysis to the three age groups targeted in both rounds (59, 63 and 67 years, Figure 1). This suggests that the decrease in sensitivity between the screening rounds was due to the older age structure alone.

Some alternative screening algorithms can also be evaluated, based on number of screen-positive findings. Had a cutoff limit of 3 ng ml^{-1} been used, the number of test-positive men would have been increased from 1980 (9.5%) to 2762 (13.3%) in the first screening round. For the second screening round, the number of screen-positive tests with a cutoff level of 3 ng ml^{-1} would have increased from 2303 to 3401 compared with the current protocol (12.3% *vs* 18.3% screen positive).

Age-specific cutoff levels (3.5 ng ml^{-1} for ages 55–59 and 4.5 ng ml^{-1} for 63–67) would have resulted in 287 fewer screen-positive findings (from 1980 to 1693, i.e. from 9.5 to 8.1%) in the first screening round, that is, slightly lower compared with the current screening protocol. The number of screen-positive results would have increased for men in their fifties and decreased for older men. In the second screening round with 4 years older subjects, age-specific cutoff levels would have increased the number of screen-positive findings by 557 compared with the protocol used in the trial (from 2303 to 2860 i.e. from 12.4 to 15.4%).

Table 3 Number of men by screening result and prostate cancer diagnosis in the second screening round of the Finnish prostate cancer screening trial

	Prostate cancer diagnosis		
Screening result	**Yes**	**No**	**Total**
Positive	583[a]	1573	2156[b]
Negative	45[c]	16 264	16 309
Total	628	17 837	18 465[b]

[a]Screen-detected prostate cancer. [b]A total of 147 screen-positive men failed to undergo biopsy and were excluded. [c]Interval cancer among screen-negative men.

Table 4 The frequency of TN and FP screening findings by age in the second screening round of the Finnish prostate cancer screening trial

Age	TN[a]	FP[b]	Specificity (95% CI)[c]
59	5700	322	0.947 (0.941–0.952)
63	4464	434	0.911 (0.903–0.919)
67	3426	395	0.897 (0.887–0.906)
71	2719	422	0.866 (0.854–0.878)
Total	16 309	1573	0.912 (0.908–0.916)

Abbreviations: CI = confidence interval; FP = false positive; TN = true negative. [a]TN: No. of men with negative screening result (serum PSA < 3.0 ng ml^{-1} or PSA 3.0–3.9 ng ml^{-1} with a negative ancillary examination (benign finding at digital rectal examination or free/total PSA ratio ⩾ 0.16)). [b]FP: No. of men with positive screening result (serum PSA < 4.0 ng ml^{-1} or PSA 3.0–3.9 ng ml^{-1} with a positive ancillary examination (suspicious finding at digital rectal examination or free/total PSA ratio < 0.16)) minus number of screen-detected cancers (147 men without biopsy excluded). [c]Specificity: TN/(TN+FP).

DISCUSSION

We report a systematic assessment of specificity in relation to its several possible determinants in a population-based trial. Our results show that a reasonably high specificity (above 90%) can be achieved with the PSA test in prostate cancer screening. Moreover, specificity decreases only slightly at repeat (incidence) screening, and this is entirely attributable to ageing of the study subjects.

Overall, specificity of serum PSA as screening test for prostate cancer was slightly above 90%. A Canadian screening study with a cutoff of 3 ng ml^{-1} reported 90% specificity (Labrie *et al*, 1992) and similar findings were reported from the US (Mettlin *et al*, 1994). A volunteer-based study in the US reported specificity of 73% (Punglia *et al*, 2005). A meta-analysis estimated specificity of PSA as 93% at 4.0 ng ml^{-1} (Mistry and Cable, 2003).

Our study population may represent relatively low-risk men, as the trial is population-based and the subjects are fairly young. Yet, the incidence of prostate cancer in Finland is rather high in international comparison, with age-standardised incidence of 84 per 100 000 in 2002 (Ferlay *et al*, 2004). Owing to the representative study population, our findings are likely to be more applicable to the general population than those from volunteer-based studies. Furthermore, we used a consistent definition of specificity, with systematic evaluation of various factors affecting specificity within the screening trial.

Specificity was only slightly lower in the second screening round compared with the first. This was due to participants being older at the second round. The main factor is probably the strong increase in prevalence of benign prostatic hyperplasia with age. Introduction of a new biopsy regimen with increased number of cores may have also decreased the number of apparent FP screening findings (if a larger proportion of true-positive findings were detected). In both rounds, the specificity was higher in the young age groups. This finding indicates that specificity is likely to decrease at subsequent screening rounds, as age at screening increases.

Digital rectal examination as an ancillary test among men with intermediate PSA levels was associated with a lower rate of FP findings than F/T PSA and hence, slightly higher specificity. The yield was also lower than with free PSA (2.1% *vs* 5.2% of men with PSA 3.0–3.9 ng ml^{-1}). This is consistent with the findings from a Dutch screening trial, where the specificity of DRE was 91% (Schröder *et al*, 1998). However, the costs for a DRE are

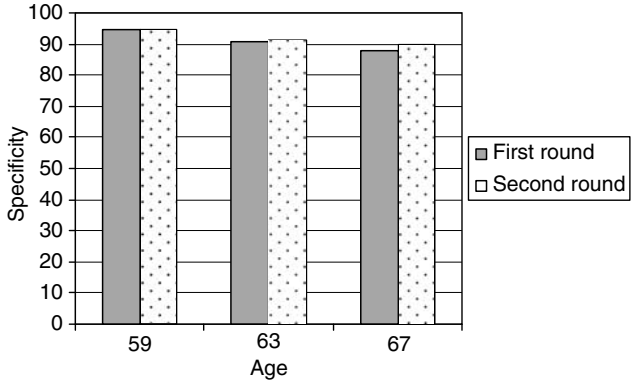

Figure 1 Specificity by age and screening round.

substantially higher than determination of F/T PSA in our trial, where a blood sample is drawn initially and can be used for determination of both total and free PSA, whereas DRE requires a separate visit for an urologist.

We estimated the specificity first by assuming that the proportion of false negatives (cancers among SN men surfacing during the screening interval) is negligible and can be ignored. This cross-sectional approach gives a measure that can be called relative specificity. Longitudinal analysis with correction for false-negative results (interval cases) is able to take into account the fact that many men with a negative biopsy do in fact harbour a latent cancer. Yet, adjustment for this did not materially affect the results. However, if all men harbouring a focal carcinoma in their prostates were classified as false negative, the situation would change dramatically as this has been very common in studies based on autopsy (Breslow *et al*, 1977; Kabalin *et al*, 1989) and cystoprostatectomy specimens or prostate tissue removed in transurethral prostatectomy (Montie *et al*, 1989; Merrill and Wiggins, 2002). Studies based on natural history models have estimated that up to 45% of screen-detected cases may be due to overdiagnosis, that is, cancers that would not have surfaced clinically during the man's lifetime if unscreened (Etzioni *et al*, 2002; Draisma *et al*, 2003). Thus, latent or minimal disease is very frequent, and there are good grounds to argue that presence of malignant histological features alone does not constitute a true golden standard for clinically significant prostate cancer. This issue can also been as a problem of FP findings, if overdiagnosed cases (if identifiable) were to be classified as FP findings. Yet, they cannot be reliably identified by current means, even if the above argument was accepted. Both issues, however, emphasise the need for definition of diagnosis of prostate cancer. We have used the conventional approach, but taking into the above uncertainties would have reduced the estimates of specificity.

Not all men with screen-positive result attend diagnostic examinations, and the results may not be available, if medical care is sought outside the screening organisation. In our material, approximately 0.5% of all participants or 5% of screen-positive men did not undergo biopsy within the trial (in the study hospitals). In the screening programme, these men are classified as negatives, that is, no further procedures are undertaken (despite indications being fulfilled). This is problematic when evaluating a screening test. In calculation of specificity, these men were assumed to be true positives and FPs in the same proportion as those biopsied. Owing to the small number of such cases, this did not affect our estimate of test specificity.

No consensus has been established as to the optimal use of PSA and several approaches have been proposed, including age-specific cutoffs and PSA relative to prostate volume (Gretzer and Partin 2003). Cutoff values even lower than 4 ng ml^{-1} have been proposed and are being used in some screening projects (Labrie *et al*, 1992; Krumholtz *et al*, 2002; Punglia *et al*, 2005). In the European Randomized Study of Screening for Prostate Cancer, ERSPC, a cutoff level of 3 ng ml^{-1} instead of 4 ng ml^{-1} was associated with increase in the proportion of test-positive findings from 1.6 to 5.1% (de Koning *et al*, 2002). Generally, both the proportion of screening-positive findings and detection rates have been higher in studies with combined modality screening (e.g., DRE and/or TRUS in addition to PSA). In our study, a limit of 3 ng ml^{-1} would have resulted in an increase in FP tests by more than a third. As the increase in screen-positive findings would be in the low PSA range, where prostate cancer prevalence is likely to be low and FP results more common than at higher PSA levels, adopting a lower cutoff level is likely to reduce specificity.

Age-specific cutoff values have been proposed for PSA in order to improve specificity of the test among older men (Oesterling *et al*, 1993). The rationale is that the prostate volume and prevalence of benign prostatic hyperplasia increase rapidly after 60 years of age. Use of age-specific cutoff levels would have resulted in a similar number of screen-positive findings in the first round, but substantially higher numbers in the second screening round. As no referrals or biopsy decision were made based on the age-specific cutoff values, we were not able to directly assess the possible effect on specificity. It would have resulted in large numbers of screen-positive men in older age groups and lower numbers in younger age groups. Because specificity was inversely correlated with age, it is likely that use of age-specific cutoff values would have resulted in lower specificity.

There are two approaches for avoiding information bias owing to PSA-driven biopsy in assessment of validity of the PSA test. First, it can be argued that everybody should receive the diagnostic test (prostate biopsy) when evaluating specificity, in order to completely identify those with disease. In some studies, all men have been biopsied, regardless of PSA result, which has resulted in detection of prostate cancer even at low PSA levels (Labrie *et al*, 1992; Thompson *et al*, 2004). These studies have also shown similar specificity for PSA as others (90–94%). Alternatively, the distortion from 'affirming the consequent' can be avoided, when no test results are followed by diagnostic examination (Walter, 1999). In serum bank studies, the PSA has been determined only afterwards and therefore it has not affected the diagnosis (Gann *et al*, 1995; Hakama *et al*, 2001). Specificity in this context has been estimated as 91–94%. Furthermore, cases in the serum bank studies have been diagnosed mainly before the PSA screening era and also therefore likely to avoid overdiagnosis.

In comparison with screening for other cancers, our results indicate similar or slightly lower specificity for PSA in prostate cancer screening. In mammography screening for breast cancer, specificity has ranged 82–99%, being commonly slightly above 90% (Elmore *et al*, 2005). Fairly similar figures (86–100%) have been reported for the cervical smear in cervix cancer screening (Nanda *et al*, 2000; Cervix cancer screening 2005). In faecal occult blood testing for colorectal cancer, slightly higher specificity (95% or higher) has been found (Allison *et al*, 1996; Rozen *et al*, 2000).

We conclude that screening for prostate cancer based on PSA determination has acceptable specificity. It should, however, be further improved if such screening is to be adopted as public health policy. We do not recommend PSA screening before the results in terms of mortality from prostate cancer are known.

ACKNOWLEDGEMENTS

The study was supported financially by the Academy of Finland (Grant no. 205862), Cancer Society of Finland, Tampere University Hospital Research Fund and Sigrid Jusélius Foundation. We are grateful to Pirkanmaa Cancer Society and Cancer Society of Finland clinics and their staff for drawing blood samples for the study, Mr Jarkko Koivuniemi for programming, and Ms Minna Heikkilä for data entry.

REFERENCES

Ablin RJ, Bronson P, Soanes WA, Vitebsky E (1970) Tissue- and species-specific antigens of normal human prostatic tissue. *J Immunol* **104:** 1329–1339

Allison JE, Tekawa IS, Ransom LJ, Adrain AL (1996) A comparison of fecal occult-blood tests for colorectal-cancer screening. *N Engl J Med* **334:** 155–159

Auvinen A, Alexander FE, de Koning HJ, Miller AB (2002) Should we start population screening for prostate cancer? Randomised trials are still needed. *Int J Cancer* **97:** 377 – 378

Breslow N, Chan CW, Dhom G, Drury RA, Franks LM, Gellei B, Lee YS, Lundberg S, Sparke B, Sternby NH, Tulinius H (1977) Latent carcinoma of prostate at autopsy in seven areas. *Int J Cancer* **20:** 680 – 688

Cervix cancer screening (2005) *IARC Handbooks of Cancer Prevention* Vol 10 Lyon: International Agency for Research on Cancer

de Koning HJ, Auvinen A, Berenguer Sanchez A, Calais da Silva F, Ciatto F, Denis L, Gohagan J, Hakama M, Hugosson J, Kranse R, Nelen V, Prorok P, Schröder FH (2002) A large-scale randomized prostate cancer screening trials: program performances in the European Randomized Screening for Prostate Cancer trial and the Prostate, Lung, Colorectal and Ovary cancer trial. *Int J Cancer* **97:** 237 – 244

Draisma G, Boer R, Otto S, Van Cruijsen IW, Damhuis RA, Schröder FH, de Koning HJ (2003) Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. *J Natl Cancer Inst* **95:** 868 – 878

Elmore JG, Armstrong K, Lehman CD, Fletcher SW (2005) Screening for breast cancer. *JAMA* **293:** 1245 – 1256

Etzioni R, Penson DF, Legler JM, di Tommaso D, Boer R, Gann PH, Feuer EJ (2002) Overdiagnosis due to prostate-specific antigen screening: lessons from US prostate cancer incidence trends. *J Natl Cancer Inst* **94:** 981 – 990

Ferlay J, Bray F, Pisani P, Parkin DM (2004) GLOBOCAN 2002: Cancer Incidence, Mortality and Prevalence Worldwide. IARC Cancer Base No. 5. version 2.0, IARC Press, Lyon, 2004 (also available at http://www-dep-iarc.fr)

Gann PH, Hennekens CH, Stampfer MJ (1995) A prospective evaluation of plasma prostate-specific antigen for detection of prostatic cancer. *JAMA* **273:** 289 – 294

Gretzer MB, Partin AW (2003) PSA markers in prostate cancer detection. *Urol Clin North Am* **30:** 677 – 686

Hakama M (1991) Screening. In *Oxford Textbook of Public Health* Holland WW, Detels R, Knox G (Eds) pp 91 – 106. Oxford: Oxford University Press

Hakama M, Stenman UH, Aromaa A, Leinonen J, Hakulinen T, Knekt P (2001) Validity of the prostate specific antigen test for prostate cancer screening: follow-up study with a abnk of 21,000 sera in Finland. *J Urol* **166:** 2189 – 2191

Hankey BF, Feuer E, Clegg LX, Hayes RB, Legler JM, Prorok PC, Ries LA, Merrill RM, Kaplas RS (1999) Cancer surveillance series: interpreting trends in prostate cancer – part I: evidence of the effects of screening in recent prostate cancer incidence, mortality, and survival rates. *J Natl Cancer Inst* **91:** 1017 – 1024

Kabalin JN, McNeal JE, Price HM Freiha FS, Stamey TA (1989) Unsuspected adenocarcinoma of the prostate in patients undergoing cystoprostatectomy for other causes: incidence, histology and morphologic observations. *J Urol* **141:** 1091 – 1094

Krumholtz JS, Carvalhal GF, Ramos CG, Smith DS, Thorson P, Yan Y, Humphrey PA, Roehl KA, Catalona WJ (2002) Prostate-specific antigen cutoff of 2.6 ng ml^{-1} for prostate cancer screening is associated with favorable pathologic tumor features. *Urology* **60:** 469 – 473

Labrie F, Dupont A, Suburu R, Cusan L, Tremblay M, Gomez JL, Emond J (1992) Serum prostate specific antigen as pre-screening test for prostate cancer. *J Urol* **147:** 846 – 851

Li TS, Beling CG (1973) Isolation and characterization of two specific antigens of human seminal plasma. *Fertil Steril* **24:** 134 – 144

Merrill RM, Wiggins CL (2002) Incidential detection of population-based prostate cancer incidence rates through transurethral resection of the prostate. *Urol Oncol* **7:** 213 – 219

Mettlin C, Littrup PJ, Kane RA, Murphy GP, Lee F, Chesley A, Badalament R, Mostofi FK (1994) Relative sensitivity and specificity of serum prostate specific antigen (PSA) level compared with age-referenced PSA, PSA density, and PSA change. *Cancer* **74:** 1615 – 1620

Mistry K, Cable G (2003) Meta-analysis of prostate-specific antigen and digital rectal examination as screening tests for prostatic carcinoma. *J Am Board Fam Pract* **16:** 95 – 101

Montie JE, Wood DP, Pontes JE, Boeytt JM, Levin HS (1989) Adenocarcinoma of the prostate in cystoprostatectomy specimens removed for bladder cancer. *Cancer* **63:** 381 – 385

Nanda K, McCrory DC, Myers ER, Bastian LA, Hasselblad V, Hickey JD, Matchar DB (2000) Accuracy of the Papanicolau test in screening for and follow-up or cervical cytologic abnormalities: a systematic review. *Ann Intern Med* **132:** 810 – 819

Oesterling JE, Jacobsen SJ, Chute CG, Guess HA, Girman CJ, Panser LA, Lieber MM (1993) Serum prostate-specific antigen in a community-based population of healthy men. *JAMA* **270:** 860 – 864

Parkin DM, Whelan SL, Ferlay J, Teppo L, Thomas DB (2002) *Cancer Incidence in Five Continents* Vol 10 Lyon: International Agency for Research on Cancer

Punglia RS, D'Amico AV, Catalona WJ, Roehl KA, Kuntz KM (2005) Effect of verification bias on screening for prostate cancer by measurement of prostate-specific antigen. *N Engl J Med* **349:** 335 – 342

Rozen P, Knaani J, Samuel Z (2000) Comparative screening with a sensitive guaiac and specific immunochemical occult blood test in an endoscopic study. *Cancer* **89:** 46 – 52

Schröder FH, van der Maas PJ, Beemsterboer P, Kruger AB, Hoedemaker R, Rietbergen J, Kranse R (1998) Evaluation of the digital rectal examination as a screening test for prostate cancer. *J Natl Cancer Inst* **90:** 1817 – 1823

Stenman UH, Finne P, Zhang W, Leinonen J (2000) Prostate-specific antigen and other prostate cancer markers. *Urology* **56:** 893 – 898

Teppo L, Pukkala E, Lehtonen M (1994) Data quality and quality control of a population-based cancer registry. *Acta Oncol* **33:** 365 – 369

Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, Minasian LM, Ford SG, Lippman SM, Crawford ED, Crowley J, Coltman Jr CA (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 – 2246

Walter SD (1999) Estimation of test sensitivity and specificity when disease confiemation is limited to positive results. *Epidemiology* **10:** 67 – 72

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

Zelen M (1979) A new design for randomized clinical trials. *N Engl J Med* **300:** 1242 – 1245