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.
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

Cover design by
Juha Siro

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

Acta Universitatis Tamperensis 1247
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 640
ISSN 1456-954X
http://acta.uta.fi

Tampereen Yliopistopaino Oy – Juvenes Print
Tampere 2007
## Contents

List of original publications .................................................................................. 4

1 Introduction ........................................................................................................... 5

2 Review of the literature ......................................................................................... 6

3 Purpose of this study ............................................................................................. 21

4 Material and methods .......................................................................................... 22

5 Results .................................................................................................................. 25

  5.1 Feasibility ....................................................................................................... 25

  5.2 PSA distribution ............................................................................................. 26

  5.3 Positive predictive value and detection rate ................................................... 27

  5.4 Sensitivity and specificity ................................................................................ 30

  5.5 Biological characteristics of cancers ............................................................... 31

6 Discussion ............................................................................................................. 34

7 Summary ............................................................................................................... 40

8 Kiitokset ............................................................................................................... 42

9 References ............................................................................................................ 46

Original Publications .............................................................................................. 55
List of original publications

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


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
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 ≥ 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 –20º 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 orchietomy) 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

<table>
<thead>
<tr>
<th>Background variable</th>
<th>Intervention</th>
<th>Control</th>
<th>Total target population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Invites</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Randomized</td>
<td>Eligible at time of invitation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N (%)</td>
<td>N (%)</td>
<td>N (%)**</td>
</tr>
<tr>
<td>Year of invitation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>8,000 (39)</td>
<td>7,281 (91)</td>
<td>5,050 (63)</td>
</tr>
<tr>
<td>1997</td>
<td>8,000 (42)</td>
<td>7,659 (96)</td>
<td>5,255 (66)</td>
</tr>
<tr>
<td>1998</td>
<td>8,000 (39)</td>
<td>7,653 (96)</td>
<td>5,377 (67)</td>
</tr>
<tr>
<td>1999</td>
<td>8,000 (39)</td>
<td>7,604 (95)</td>
<td>5,111 (64)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>10,495 (40)</td>
<td>9,911 (94)</td>
<td>6,457 (62)</td>
</tr>
<tr>
<td>59</td>
<td>8,378 (40)</td>
<td>7,915 (94)</td>
<td>5,552 (66)</td>
</tr>
<tr>
<td>63</td>
<td>6,931 (40)</td>
<td>6,555 (95)</td>
<td>4,729 (68)</td>
</tr>
<tr>
<td>67</td>
<td>6,196 (40)</td>
<td>5,816 (94)</td>
<td>4,055 (65)</td>
</tr>
<tr>
<td>Area of residence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helsinki</td>
<td>24,016 (40)</td>
<td>22,601 (94)</td>
<td>14,989 (62)</td>
</tr>
<tr>
<td>Tampere</td>
<td>7,984 (39)</td>
<td>7,596 (95)</td>
<td>5,804 (73)</td>
</tr>
<tr>
<td>Total</td>
<td>32,000 (40)</td>
<td>30,197 (94)</td>
<td>20,793 (65)</td>
</tr>
</tbody>
</table>

* 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

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

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

<table>
<thead>
<tr>
<th>PSA (ng/ml)</th>
<th>No. of men</th>
<th>No. referred to ancillary test</th>
<th>No. of men biopsied</th>
<th>No. of screen-detected cancers</th>
<th>PPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>DRE only</td>
<td>Referral based on</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DRE</td>
<td>F/T</td>
<td>PSA</td>
<td></td>
</tr>
<tr>
<td>–1.99</td>
<td>15,725</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2 –2.99</td>
<td>2,171</td>
<td>119</td>
<td>7</td>
<td>119</td>
<td>7</td>
</tr>
<tr>
<td>3 –3.99</td>
<td>1,071</td>
<td>801</td>
<td>82</td>
<td>60</td>
<td>861</td>
</tr>
<tr>
<td>4 –9.99</td>
<td>1,525</td>
<td>–</td>
<td>–</td>
<td>1,525</td>
<td>1,525</td>
</tr>
<tr>
<td>10 –</td>
<td>301</td>
<td>–</td>
<td>–</td>
<td>301</td>
<td>301</td>
</tr>
<tr>
<td>Total</td>
<td>20,793</td>
<td>920</td>
<td>89</td>
<td>1,826</td>
<td>2,806</td>
</tr>
</tbody>
</table>

*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 ≥ 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

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

*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

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

<table>
<thead>
<tr>
<th>Age</th>
<th>Test negative men*</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Interval cancers</td>
<td>Person-years</td>
<td>Incidence</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>8</td>
<td>16,773</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>4</td>
<td>13,836</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>7</td>
<td>11,505</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>5</td>
<td>9,536</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>51,650</td>
<td>48**</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Control group</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Interval cancers</td>
<td>Person-years</td>
<td>Incidence</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>44,243</td>
<td>158</td>
<td></td>
</tr>
<tr>
<td>124</td>
<td>34,641</td>
<td>361</td>
<td></td>
</tr>
<tr>
<td>139</td>
<td>28,299</td>
<td>491</td>
<td></td>
</tr>
<tr>
<td>206</td>
<td>25,227</td>
<td>566</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>539</td>
<td>132,410</td>
<td>394**</td>
</tr>
</tbody>
</table>

Sensitivity

| | |
|---|---|---|---|
| | | | 0.70 |
| | | | 0.92 |
| | | | 0.88 |
| | | | 0.91 |
| | | | 0.85*** |

*PSA <3 ng/ml or PSA 3-4 ng/ml with either benign DRE finding (1996-1998) or free/total PSA ratio ≥ 0.16 (1999)

**Among screen negative men (Ii), in the control group (Ic)

***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

<table>
<thead>
<tr>
<th>Age</th>
<th>Negative test</th>
<th>Positive test, negative confirmation</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Interval cancers</td>
<td>Total</td>
</tr>
<tr>
<td>55</td>
<td>6,153</td>
<td>7</td>
<td>214</td>
</tr>
<tr>
<td>59</td>
<td>5,137</td>
<td>11</td>
<td>279</td>
</tr>
<tr>
<td>63</td>
<td>4,146</td>
<td>11</td>
<td>409</td>
</tr>
<tr>
<td>67</td>
<td>3,389</td>
<td>13</td>
<td>456</td>
</tr>
<tr>
<td>Total</td>
<td>18,825</td>
<td>42</td>
<td>1,358</td>
</tr>
</tbody>
</table>

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

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

*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

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


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 pulpahvatat ideat ovat pitäneet vauhtia yllä. Hänen osuutensa artikkelien tuottamisessa on ollut ratkaiseva. 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 tukenuut ja tuuppinut, puolustanut ja auttanut eteenpäin. Minnalta olen oppinut sen minkä tiedän seulontojen käytännön organisoinnista ja hänen sihteerintaidostaan olen yrittänyt omaksua edes murto-osan. Minnan kanssa käydyt pohdinnat ja keskustelut ovat olleet hedelmällisiä ja ne tietenkin jatkuvat yhä.

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

Professori Teuvo Tammelalla on ollut painava osuus tutkimuksen urologian asianmukaisuus, 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 seurantaryhmän 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. Häntä on myös riittänyt huomiota ja tukea väitöskirjatyön etenemiselle, johon toimi Kansanterveyden tutkijakoulun johtajana antoi lisäpontta.


PSA-näytteet otetaan Helsingissä Bulevardin Erikoislääkärin, Syöpäjärjestöjen poliklinikalla ja Tampereella Pirkanmaa 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 Biomedical Helsingissä, professori Ulf-Håkan Stenmanin johdolla. Muutosta huolimatta työntekijät ovat pysyneet samoina ja olenkin suuren kiitoksen velkaa laboranttii Maarit Leinimaalle ja laboratorioihoitaja 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 sairaalissa, 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 mitaan useammatkin nuoret ja iäkkäimmät henkilöt lyhyemmän tai pidemmän jakson. Kaikki on kuitenkin tallessa ja siitä heille kiitos.

Kiiitän esitarkastajani professori Olavi Lukkarista ja emeritusprofessori Mauri Isokoskea heidän työstäni antamista komenteista ja tuiki tarpeellisten täydennysten neuvomisesta.

Koko PSA-tutkimusryhmä, joka on laajentunut toiminnan vuosien mittaan useammatkin nuoret ja iäkkämmät henkilöt lyhyemmin tai pidemmin jakson. Kaikki on kuitenkin tallessa ja siitä heille kiitos.

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

Koko PSA-tutkimusryhmä, joka on laajentunut toiminnan vuosien mittaan useammatkin nuoret ja iäkkämmät henkilöt lyhyemmin tai pidemmin jakson. Kaikki on kuitenkin tallessa ja siitä heille kiitos.

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

Koko PSA-tutkimusryhmä, joka on laajentunut toiminnan vuosien mittaan useammatkin nuoret ja iäkkämmät henkilöt lyhyemmin tai pidemmin jakson. Kaikki on kuitenkin tallessa ja siitä heille kiitos.

Kiitän esitarkastajani professori Olavi Lukkarista ja emeritusprofessori Mauri Isokoskea heidän työstäni antamista kommenteista ja tuiki tarpeellisten täydennysten neuvomisesta.
Ystäväni, varsinkin jengiläiset, joiden kanssa olen maailmaa parantanut eri yhteyksissä jo 45 vuotta, ovat kiitoksen ansainneet. He ovat jaksaneet uskoa, että tämänkin opiskelujakso elämässäni joskus päätyy positiiviseen tulokseen.


Helsingissä elokuussa 2007

Liisa Määttänen
9 References


Original Publications
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...
examination in 1996.

ng/mL; in addition, 108 men with a PSA concentration of 2.0

Specificity and acceptability of a prostate cancer screen-
ting program. To increase specificity, we 

The high participation rate (69%) indicates that, in Finland, prostate cancer screening will be feasible, if it is found to be effective. At baseline, opportunis-
tic screening was relatively low in Fin-
land and does not seem to jeopardize the 

We found a PPV for serum PSA con-
detected cancer at higher cutoff levels. 

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

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

*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.

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

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

*Percentage of participants.

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

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.

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 Heikkila for data management.

Manuscript received September 18, 2000; revised November 8, 2000; accepted January 23, 2001.
Specificity of serum prostate-specific antigen determination in the Finnish prostate cancer screening trial

L Määttänen¹,² M Hakama¹,², TLJ Tammela³, M Ruutu⁴, M Ala-Opas⁵, H Juusela⁵, P Martikainen⁶, U-H Stenman⁷ and A Auvinen*,²,⁸
¹Finnish Cancer Registry, Lisankatu 21 B, FIN-00170 Helsinki, Finland; ²Tampere School of Public Health, FIN-33014 University of Tampere, Tampere, Finland; ³Department of Urology, Tampere University Hospital and University of Tampere, Box 2000, FIN-33521 Tampere, Finland; ⁴Department of Urology, Helsinki University Hospital, Box 580 FIN-00029, Helsinki, Finland; ⁵Department of Surgery, Jorvi Hospital, Turuntie 150, FIN-02740 Espoo, Finland; ⁶Department of Pathology, Tampere University Hospital, University of Tampere, Box 2000, FIN-33521 Tampere, Finland; ⁷Department of Clinical Chemistry, Helsinki University Hospital, Box 700, FIN-00029 Helsinki, Finland; ⁸Finnish Cancer Institute, Lisankatu 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 and FP results (SN/(SN + FP)). A SN finding was defined as either PSA ≤ 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 (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.


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
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 80458 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 >4 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 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 ng ml\(^{-1}\) or PSA 3.0–3.9 ng ml\(^{-1}\) in combination with negative DRE or free/total ratio \(\geq 0.16\) (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).

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 (102/1866) 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 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 1414 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 increase in the number of SN findings.

**RESULTS**

In the target population of 80458 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, 30194 were eligible and invited. Of them, 20794 (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, 18825 had serum PSA concentration below 3 ng ml\(^{-1}\) or PSA 3.0–3.9 ng ml\(^{-1}\) in combination with negative DRE or free/total ratio \(\geq 0.16\).

<table>
<thead>
<tr>
<th>Screening result</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>508</td>
<td>1358</td>
<td>1866</td>
</tr>
<tr>
<td>Negative</td>
<td>42</td>
<td>18 783</td>
<td>18 825</td>
</tr>
<tr>
<td>Total</td>
<td>550</td>
<td>20 141</td>
<td>20 691</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Age</th>
<th>TN(^a)</th>
<th>FP(^b)</th>
<th>Specificity (95% CI)(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>6153</td>
<td>214</td>
<td>0.966 (0.962–0.971)</td>
</tr>
<tr>
<td>59</td>
<td>5137</td>
<td>279</td>
<td>0.948 (0.943–0.954)</td>
</tr>
<tr>
<td>63</td>
<td>4146</td>
<td>409</td>
<td>0.910 (0.902–0.919)</td>
</tr>
<tr>
<td>67</td>
<td>3389</td>
<td>456</td>
<td>0.881 (0.871–0.892)</td>
</tr>
<tr>
<td>Total</td>
<td>18 825</td>
<td>1358</td>
<td>0.933 (0.929–0.936)</td>
</tr>
</tbody>
</table>

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

Abbreviations: CI = confidence interval; FP = false positive; TN = true negative. \(^a\)TN: No. of men with negative screening result (serum PSA <4.0 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. 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 2,303 screen-positive subjects referred to biopsy. Of them, 2,156 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%).

**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...
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 a 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 (Greter 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 larger 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 Juselius Foundation. We are grateful to Pirkkannma 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 Heikilä for data entry.

REFERENCES


Clinical Studies


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