RESEARCH PLAN: ProScreen (Prostate Cancer Screening Trial)

1. INVESTIGATORS

Primary investigator:	Anssi Auvinen, professor of epidemiology, School of Health Sciences, University of Tampere, Tampere, Finland
Co-PI:	Antti Rannikko, professor of urology, Dept. of Urology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

2. BACKGROUND

Screening has intuitive appeal as means for disease control, as the promise of detecting disease cases early appears to imply improved potential for cure and superior treatment outcomes. Yet, these are several misconceptions that can lead to overestimation of the screening benefits, unless rigorous study design is used, such as a randomised trial (Auvinen & Hakama 2016). Simply detecting disease does not indicate beneficial effect of screening, but it should decrease either mortality or incidence of the target disease (Auvinen&Hakama 2016). Screening aims to detect cancer early and thereby allow curative treatment of cases that would otherwise have been diagnosed at an advanced stage, with poorer treatment outcomes. The spectrum of disease detectable among asymptomatic subjects can be very different from those in the clinical realm. Biological characteristics of latent disease may be dramatically different from the cases presenting in the symptomatic stage, and this can mean entirely different disease course and natural history. However, just showing that cases are detected through screening and that they possess features suggesting favourable prognosis is not valid indication of screening benefit. First, screening has more potential to detect slow-growing cases than more aggressive tumours (length bias), as the former have a longer duration of the detectable pre-clinical phase (sojourn time). Second, screening can also detect tumours when they are small, without affecting the treatment outcome i.e. lead to earlier diagnosis without affecting the time of death (lead-time bias). Third, screening can lead to detection of indolent disease that would have remained latent (clinically silent) in the absence of screening (overdiagnosis). Therefore, mortality reduction in a randomised trial is the gold standard for screening evaluation (Auvinen & Hakama 2016)

The potential benefit of a therapeutic intervention for clearly progressive disease normally encountered in the clinical setting is not readily applicable when managing disease detected at an early, asymptomatic stage by screening (Welch & Black 2010). The latter encompass a substantial proportion of cases with halted or exceedingly slow progression (Esserman 2014). In prostate cancer, there is substantial reservoir of latent disease, with a prevalence of 20% at age 50, approaching 30% by age 60 (Bell 2015, Jahn 2015). Detection of these cases will involve no benefit for the man, but still carries all the risks and adverse effects. This would not be an issue, if it was possible to disentangle the spectrum of disease with well-established prognostic predictors so that indolent disease could be left untreated. However, we lack such indicators for prostate cancer. The Gleason score remains the best prognostic indicator, and a Gleason sum <7 is

regarded as an indication of disease that has only questionable clinical importance (Leapman 2017, Epstein 2016).

Prostate cancer (PrCa) is currently one of the key medical and public health challenge: It is the most common cancer among men in Western countries with incidence around 90 per 100,000 in Western Europe. Further, it is the third most common cause of cancer death in men in Europe (GloboCan 2012). In Finland, incidence has decreased from 120 to 80 in the 2000's. The prostate cancer epidemic is, however, largely iatrogenic, as the increase in incidence is attributable predominantly to increased serum prostate-specific antigen (PSA) testing.

Overdiagnosis of prostate cancer

A substantial proportion of the currently diagnosed prostate cancer cases represent overdiagnosis. Overdiagnosis is defined as detection of disease that would not have caused any harm during a man's lifetime, i.e. unnecessary detection of latent disease of dubious clinical relevance. The mortality impact of screening is achieved by identifying potentially lethal cancers and halting their progression through curative treatment. By definition, overdiagnosed cases cannot contribute to the mortality reduction achievable by screening. Hence, avoiding them does not diminish the screening benefits. For a man with non-progressive cancer, treatment involves adverse effects of treatment (e.g. erectile dysfunction and urinary incontinence) without benefit to the patient. At the population level, the fact that relative survival of local prostate cancer at five years can exceed 100% (McPhail 2015, Seikkula 2017) indicates lack of excess mortality, inconsistent with the behaviour of a truly malignant disease.

Overdiagnosis is more extensive in prostate cancer screening than other cancer screening programmes: Modelling studies have estimated that it amounts to 40% of the screen-detected cases (Draisma 2009), amounting to three excess cases per 100 men invited to screening (Wu 2012) or 40% of screen-detected tumours (Draisma 2009), while for mammography screening the excess of breast cancer has ranged 10-30% (Falk 2013, Beckmann 2015, Welch 2016). Hence, the challenge is in detecting only those cases that benefit from treatment, with real potential for progressing to overt clinical disease.

A further complication for management of early prostate cancer is the lack of minimally invasive therapeutic approaches for low-risk disease. In breast cancer, breast conserving surgery (lumpectomy or partial mastectomy) can be used for early, screen-detected disease, but in prostate cancer, no effective focal therapy has been established, and the only curative treatment modalities are surgery (radical prostatectomy, currently often robot-assisted) and radiotherapy (external beam, often intensity-modulated, or brachytherapy). They can achieve favourable treatment outcomes, but involve also frequent adverse effects including erectile dysfunction, urinary incontinence or proctitis (Barocas 2017, Chen 2017, Donovan 2016). For low-risk disease, active surveillance results in very high survival with lower rates of adverse effects, but refraining from active treatment for screen-detected disease does not concur with the idea of screening (avert unfavourable disease outcomes by intervening early in the disease course) – detecting disease that is not treated cannot reduce disease burden. This underscores the importance of avoiding overdiagnosis.

How can we predict disease course in prostate cancer

The Gleason score remains the best single prognostic predictor in prostate cancer. Grading in prostate cancer is based on the Gleason system (rather than WHO), obtained as the sum of the most common and second most common patterns of differentiation (tissue architecture), both assessed on a scale 1-5 (tertiary pattern also considered for biopsy-based scoring). The classification has been revised several times.

A substantial difference in outcome between the 3+4 and 4+3 has been demonstrated, with HR approximately 2.5 for 3+4 vs. 5.5 for 4+3 relative to GS<7, and 5-year biochemical relapse free survival 82-

83% vs 65-74% (Pierorazio 2013, Epstein 2016b). Recently, a prognostic grade grouping was proposed combining Gleason<7 (currently, scores below 6 are very rarely assigned), while dividing Gleason 7 into 3+4 and 4+3 based on the predominant pattern (Pierorazio 2013, Epstein 2016a). With the current grading system, Gleason grade 2-6 tumours have such a low risk of progression (5-year recurrence free rates 89-95%, Pierorazio 2016, Loeb 2016) that they can be regarded as clinically insignificant, and there is even debate about whether these tumours should be called cancers (e.g. Carter 2012), or something else (e.g. indolent lesion of epithelial origin). Hence, for pragmatic purposes, Gleason<7 tumours can be used as a simple and robust measure of overdiagnosis.

What has been achieved with prostate cancer screening?

Screening offers potentially a means for reducing prostate cancer mortality. We have shown in the European randomised screening trial (ERSPC) a 20% reduction in prostate cancer mortality at 13 years of follow-up (Schröder 2014). The absolute screening effect has increased with duration of follow-up, but it remains modest, at an absolute risk reduction of 0.1 deaths per 1000 man-years (or one death per 781 men invited to screening) (Schröder 2014, Auvinen 2016). An analysis of 16-year data is on-going and preliminary findings indicate a similar relative effect and still increasing absolute effect (Hugosson et al., in preparation). The lack of any mortality reduction in the U.S. PLCO trial has been shown to reflect high contamination and low biopsy compliance, so that the apparently conflicting results between the ERSPC and PLCO studies can be explained by shortcomings in the conduct of the U.S. trial and hence the European trial better reflects the true effect that can be achieved by screening (Gulati 2012, Palma 2016, Tsodikov submitted).

The mortality reduction achieved with PSA-based screening is comparable to the well-established cancer screening modalities i.e. mammography screening for breast cancer and faecal occult blood test for colorectal cancer (Beral 2011, Lin 2016).

The balance of benefits and harms for prostate cancer screening remains, however, debatable, with varying estimates of overall quality of life impact and cost-effectiveness (Heijnsdijk 2012, 2014, Lao 2015) and the U.S. PSTF made a recommendation against prostate cancer screening (grade D) in 2014, though a revised draft (US PSTF 2016) suggests individual decision making for men aged 55-69 with small potential benefit and harm due to overdiagnosis and false positive results. It is the adverse effects that tip the balance against prostate cancer screening, and therefore we should focus on reducing them.

An analysis of the ERSPC results by centre (Auvinen 2016) showed that the extent of benefit (mortality impact) in any centre is directly proportional to the amount of harms (overdiagnosis). This suggest that as long as PSA-based screening is used, increased effectiveness means more adverse effects – and conversely, minimising overdiagnosis would likely results in reduction of the mortality impact. This casts doubt on the possibilities of improving screening outcomes as long as it is based on PSA alone.

Hence, we must pursue either improved sensitivity (larger screening effect), or increased specificity for clinically relevant cancer (better specificity), or both. Our argument is that with recent developments in understanding of prostate cancer and advances of its detection methods, we can and should focus our efforts on decreasing overdiagnosis (reducing harms to improve the net effect).

Targeted screening is not the answer

One option for improving the balance of benefits and harms is risk group screening, i.e. targeting high-risk groups rather than entire population, or tailoring screening according to risk level (‘personalised’ or ‘risk-adaptive’ screening). Yet, accurate prediction of the risk clinically relevant prostate cancer has turned out to be highly challenging.

We collected extensive genetic data in the FinRSPC trial and have participated in the international consortia PRACTICAL, iCOGS and OncoArray. Even though these analyses have shown that prostate cancer that >100 SNPs are involved in prostate cancer aetiology (Eeles 2013, Al-Olama 2014, Dadaev submitted), polygenic prediction models have shown only moderate performance in predicting prostate cancer risk (AUC 0.67 based on 65 SNPs) with little additional predictive capacity beyond PSA (Szulkin 2015). Furthermore, most genetic risk markers tag low-risk prostate cancer better than high-risk prostate cancer and conflicting results have been obtained in different populations, suggesting heterogeneous involvement of genes across populations (population stratification). Hence, risk prediction using genetic factors has not so far allowed sufficient basis for targeting screening.

Also, even if family history is one of the strongest and best-established risk factors, our findings in the FinRSPC showed no advantage by targeting men with family history (at least one affected first-degree relative) in terms of PSA test performance or mortality reduction (Saarimäki 2015). Similar results have also been reported elsewhere (Randazzo 2016). A polygenic risk score based on extensive genotyping was correlated with the extent of overdiagnosis (Pashayan 2015). Among men with a polygenic risk score above the median, the proportion of overdiagnosed cases was lower than in the others, but the absolute risk was nevertheless higher. A more limited polygenic risk score, with 48 SNPs selected as the most informative using Finnish population data gave a low AUC (<0.6) (Sipeky, manuscript).

Further, the public health impact of screening focusing only some subgroups of the population will necessarily be limited, due to the fact that the population attributable risk (proportion of the entire disease burden) remains modest in any high-risk sub-group, as in reality the excess risk does not compensate for the small population size, as out of all people who will be diagnosed in the population, only a minor fraction to the high-risk group (for instance, if 2% of the men have a 10-fold risk, reducing prostate cancer mortality by 20% in them would not result in more than 4% reduction in prostate cancer deaths in the entire population).

Which screening test should be used?

Even though PSA is one of the best cancer biomarkers developed, it has turned out to be insufficient as a stand-alone test for prostate cancer screening, due to low specificity for clinically relevant cancer. The best option for improving performance of PSA as a screening test is combine it with other biochemical indicators.

A panel of four kallikreins (total PSA, free PSA, intact PSA and human kallikrein-related peptidase-2, hK2), known as the 4Kscore has been developed to indicate the probability of an aggressive prostate cancer. It has been evaluated prospectively in a large multi-institutional study in the US (Parekh 2015) and was recently shown to improve the predictive power of PSA alone from AUC 0.65 to 0.75 in the FinRSPC (Assel et al, submitted), with similar results reported elsewhere (Carlsson 2013, Braun 2016). With a cut-point of 6% risk for Gleason score (GS) 7+ cancer, it reduced the number of biopsies by 43%, while detecting 89% of GS≥7 cases (Bryant 2015). Prediagnostic 4Kscore is also predictive of prognosis in prostate cancer (Stattin 2015). Yet, it has never been used for screening in a randomised trial.

Optimising the diagnostic process for clinically relevant cancer

Multiparametric magnetic resonance imaging (mpMRI) has been shown to decrease the number of biopsies by 20-54%, with a substantial reduction in detection of GS<7 cancer (40%), but a very high sensitivity (90-93%) for GS 7+ prostate cancer (Ahmed 2017, Alberts 2016, Fütterer 2016). The positive predictive value has ranged 38-66% (Fütterer 2016). The recent PROMIS trial using the latest technique and procedures with 576 men showed sensitivity of 93%, with a positive predictive value of 51% and a negative predictive value of 89% (Ahmed 2017). It has been proposed that the MRI does not detect tumour tissue representing Gleason pattern 3, but only Gleason 4-5. Hence, MRI can reduce both the number of biopsied men. In

addition, the number of cores per man is lower when only suspicious foci are targeted instead of systematic biopsies (and thereby the risk of biopsy complications is also reduced). A recent systematic review showed a higher detection of clinically significant cancer (sensitivity 0.91 vs 0.76) and 44% lower detection of low-risk cancer for mpMRI-based targeted biopsy than transrectal ultrasound (TRUS) -based systematic biopsy (Schoots 2015). This has rapidly transformed the landscape of prostate cancer diagnostics, so that MRI has become a new standard for detection of clinically relevant prostate cancer within the past 12 months. Directing the biopsy to the suspect lesion allows to decrease proportion of cancers missed – estimated around 25% for systematic 10-12 core biopsy (Bjurlin 2013).

Optimising the screening algorithm

Screening frequency is an important determinant of overdiagnosis (Heijnsdijk 2009). Etzioni et al. (2013) estimated that 27% of overdiagnosis could be reduced by using longer screening intervals in men with low PSA. The PLCO trial in the U.S. was unable to show a mortality reduction despite shorter screening interval than in the ERSC (annual vs four years, Pinsky 2017). The two-year vs. four-year interval does not explain the differences between the ERSPC centres (Kilpeläinen 2015, Nevalainen 2017, Saarimäki submitted). However, a very low risk clinically relevant cancer among men with PSA<1 ng/ml even after 15 years suggest that repeat screening can be postponed beyond four years (Carlsson 2014).

In the ERSPC, no systematic differences have been shown in the screening effect by age group. Actually, the largest screening effect was in the oldest age group in Finland and the Netherlands, but among the youngest men in Sweden. Therefore, no compelling evidence exists to justify major alteration of the age range from the ERSPC (with a core age group 55-69 years). Moreover, statistical power is increased in older ages, as the risk of prostate cancer death is higher. Starting at 65 allows at least two screening round by age 71 – an age by which a large majority of the men still remain eligible for major surgery. Nevertheless, men older than 70 are not always eligible for curative treatment and their life expectancy may limit the life-years that can be gained by screening.

3. TARGET POPULATION

We will identify all 67,000 men aged 50-63 years and residing in Tampere or Helsinki from the Population registry, and after exclusion of men with a prior diagnosis of prostate cancer (expected number 200), randomize them into two trial arms in ratio 1:3 (16,750 men allocated to the screening arm expected). Randomization is based on computer-generated random numbers and will be conducted in five batches annually (10% of the entire population at a time). This procedure is chosen in order to include only men alive and free of prostate cancer (with up-to-date information on cancer and vital status). Men with a permanent address abroad are not eligible.

4. METHODS

Study rationale

Frequent adverse effects have so far tipped the balance of benefits and harms against prostate cancer screening, and therefore we will focus on employing the best possible means for reducing them. The project proposal introduces a novel concept for PC screening that minimises overdiagnosis and overtreatment, while retaining the mortality benefit to shift the balance of screening benefits and harms to a favourable net effect. The strategy for implementation as a randomised screening trial utilises three levels of risk assessment (PSA, kallikrein panel and MRI) before the diagnostic procedure (prostate biopsy), each aimed at eliminating detection of indolent disease. The study hypothesis is that by virtue of the novel

three-tiered screening algorithm, the beneficial screening effect (prostate cancer mortality reduction) can be retained, while the overdiagnosis can be largely eliminated. The impact of our integrative approach has never been evaluated – each of the methods has only been assessed in isolation. The breakthrough potential of the proposal lies in combining the three novel approaches and taking them to the forefront of applied research through a randomised trial. The key impact of the study is in defining whether the overall balance of benefits and harms of prostate cancer screening can be reversed by applying the best possible methods to detect only clinically important disease. If the study hypothesis is affirmed, it opens the way to introduction of prostate cancer screening. If the balance of harms and benefits is still unfavourable, the problem of overdiagnosis in prostate cancer may be intractable.

The main adverse effect of screening, i.e. overdiagnosis has to be minimised. The best potential methods for avoiding it are: a kallikrein panel, multiparametric MRI and flexible screening interval. Each of the methods has been shown to reduce the detection of Gleason<7 prostate cancer by at least a third, while missing only 10-15% of the more aggressive cases. Their combined effect has not, however, been studied, nor have they been evaluated in a randomised screening trial, only non-randomised evaluation of each method separately has been reported.

Finland provides a unique opportunity for carrying out the trial, because of (1) the centralized public health care allowing uniform diagnostic assessment and treatment of men in both arms in the same center by the same physicians, avoiding the possibility that differences in management by trial arm could lead to differences in prostate cancer mortality; (2) the Finnish health care registers allow comprehensive follow-up to identify all prostate cancer cases and obtain information on their characteristics, treatment received and treatment outcomes; and (3) the Finnish population has a positive attitude toward research and is eager to participate in medical research.

The fundamental idea of the trial is the uncertainty about the effects of prostate cancer screening, and the rationale rests on the assumption that the approaches that have never been evaluated in a randomised trial and never combined into a single protocol can be successfully integrated into a winning formula. This makes the trial a high-risk effort, and hence suitable for ERC. Also, the scale of the trial is such it can provide compelling evidence, owing to the very large size, and this requires funding beyond national resources.

Screening protocol

Briefly, we will first screen men with total PSA and then use an array of four biochemical markers (4Kscore) among men with PSA>3 ng/ml to improve the diagnostic accuracy for clinically significant cancer (Gleason≥7), while reducing the number of biopsies and detection of Gleason<7 cancers. For diagnostic assessment of screen-positive men, we will use multiparametric MRI, aimed at reducing the detection of Gleason<7 prostate cancer by at least 50% by the combination of the screening test and diagnostics (Figure 1). Finally, the baseline serum PSA is used for determining the optimal screening interval, ranging 2-6 years.

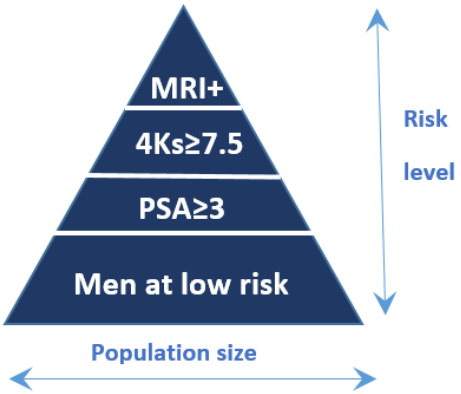

Figure 1. Using three layers of risk stratification to identify men at highest risk.

The control arm will receive usual care, with no intervention (apart from collection of registry data for all men in both arms on the outcomes, including PSA testing, prostate cancer incidence and mortality, as well as questionnaire surveys for some subgroups).

Men in the screening arm will be invited to the first screen over a two-year period. We expect similar participation as in the previous Finnish prostate cancer screening trial (two thirds or 11,167 participants). This would mean that we would screen roughly 560 men per month (assuming 10 months of effective screening annually) (Figure 2).

In the baseline survey at entry to the study, screening participants are asked about their family history of prostate cancer, as well as established or suspected risk factors for prostate cancer (including BMI, alcohol consumption, physical activity and consumption of some foodstuffs). Information will be also collected on previous PSA tests and possible prior prostate biopsy from the hospital databases.

A 30 ml blood sample will be drawn at one of the service sites of the regional laboratory networks (HUSLab and FimLab) to an EDTA tube, with plasma and serum separated and frozen. Total and free PSA are analysed at HUSLab and FimLab with electrochemiluminescence immunoassay (ECLIA) for COBAS® 6000 analyser (Roche Diagnostics, Rotkreuz, Switzerland). Calibration is performed using the WHO standards (WHO 96/670 and 68/668 for total and free PSA).

Besides screening tests, biological samples will be collected to the local biobank (a serum separation tube, a plasma separation tube, a 10 ml EDTA tube for cell-free DNA, a PAXgene tube and whole blood in two 4 ml EDTA tubes for DNA, as well as a morning urine sample). These will be used for developing prognostic models and identifying risk predictors for prostate cancer with the Prostate Cancer Research Centre at University of Tampere (prof. Tapio Visakorpi, genomic analyses) and University of Turku (Prof. Kim Pettersson, Dept. of Biotechnology, University of Turku, serum lectins and urine kallikrein speciation). Kallikrein glycovariants have shown promise as novel methods for detection of early, clinically relevant PrCa, and we will utilize lectin-assisted targets to identify cancer-associated changes in the carbohydrate moiety at Asn-45 of PSA. The rationale is based on lectins (carbohydrate binding proteins having the unique specificities for glycan structures), which are covalently coupled to fluorescent nanoparticles and used to specifically detect cancer associated glycans on PSA. From a library of lectin nanoparticles established, the most promising lectins will be identified to preferentially detect PSA from men with PrCa, whereas urine or seminal plasma derived PSA of healthy individuals remains non-reactive. Preliminary results demonstrate improved cancer specificity (improved discrimination of high grade PrCa from biopsy negative and Gleason score 6 patients) in urine using a plant lectin nanoparticle-aided PSA assay (Kekki 2017).

We will also test an ultrasensitive artificial olfaction system based on ion mobility spectrometry, a differential mobility spectrometry and a field-asymmetric ion mobility spectrometry, capable of detecting a vast range of volatile organic compounds at particle per million-trillion (ppm-ppt) concentrations in a 5 ml sample. A semi-supervised machine learning algorithm is developed to recognize a typical multidimensional pattern of volatile compounds (mainly amines) characterizing prostate (prof. Niku Oksala). In a small clinical material, an AUC of 0.77, with sensitivity 0.78 and specificity of 0.67 in distinguishing prostate cancer patients from cancer-free subjects (Roine 2014), and the current set-up has an order of magnitude lower detection threshold.

Based on the FinRSPC first round results, we expect that the proportion of men with PSA≥3 is 13% overall (10% in men aged 53-57, 17% for ages 58-62 and 24% at ages 63-65), total expected N=1452. This would mean on average 75 kallikrein panel determinations per month.

Screening interval is based on the initial PSA: men with PSA≥3 are re-invited after two years, those with PSA>1.5 after four years, and the men in the lowest risk group (PSA<1.5) after six years.

After serum PSA as the initial test, we will use a four-kallikrein panel (4Kscore) to identify men at an increased risk of a clinically relevant prostate cancer. The determinations will be performed by prof. Hans Lilja in his lab in Malmö, with samples shipped every two weeks. These men will be referred to MRI and only MRI-positive men will be biopsied (randomised to targeted biopsy alone versus targeted and systematic biopsy).

The men with the highest risk of Gleason 7+ prostate cancer based on the kallikrein panel (out of those with PSA>3) will be regarded as screen-positive and referred to diagnostic examination at Tampere or Helsinki University Hospital urology clinic. The expected frequency is 871 men in the first round (60% of men with PSA≥3, or 7.8% of all participants), on average 45 men referred to MRI each month. This is substantially lower than in the PSA-based ERSPC trial (where on average 13% of the participants were screening-positive and referred to biopsy in the first round).

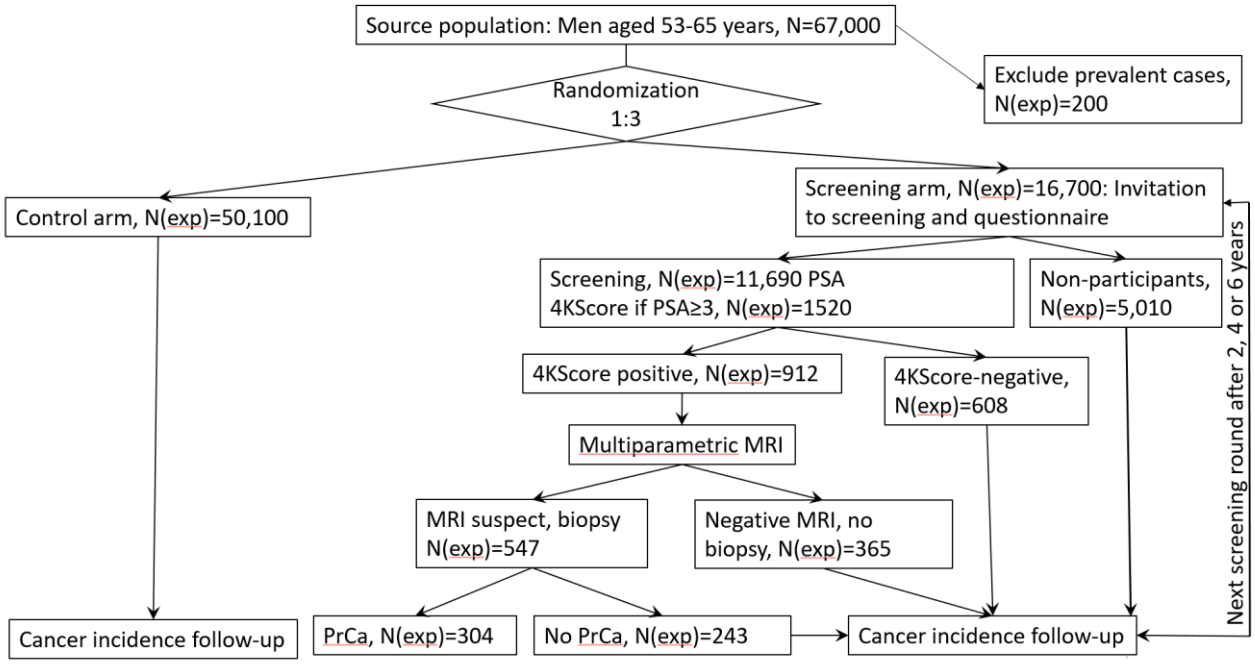

Figure 2. A CONSORT diagram illustrating the trial protocol and the expected distribution of men by test results

Diagnostic procedures

The screen-positive men will undergo a multiparametric MRI using 3T equipment (Siemens Achieva) with 32-channel pelvic phased-array coil. T2-weighted, diffusion-weighted and dynamic contrast-enhanced imaging is employed in accordance with the European Society for Urogenital Radiology guideline. The findings will be classified according to the Prostate Imaging Reporting and Data System (PIRADS-2), which is a 5-point scale to combine the MRI findings and indicate the likelihood of a significant cancer (Osses 2017). A score of 4-5 indicates a PPV>80% and also a score of 3 is regarded as a positive finding warranting biopsy. Standardisation of MRI procedures is ensured prior to the start of the trial, including slice orientation and thickness, and size of the matrix and field of view. Training sessions will be repeated annually throughout the conduct of the trial. Independent double reading for a random sample of the images is used for quality assurance, and to reduce false negative and false positive findings. Standardised and structured radiologists' reports will be obtained and prostate volume also recorded. Pseudoanonymised images are uploaded into the cloud-based trial database using the ECRF software tailored for the trial purposes.

The men with positive MRI will undergo targeted biopsy using fusion MRI-TRUS guidance (UroNav), with an additional core per each incremental 6 mm size of the lesion. The expected frequency of positive MRI is 60% of screen-positive men (4.7% of all screening participants), with a total of 521 men in the first round. To avoid missing cancers among high-risk men, those with PSA density (serum PSA divided by prostate volume) >0.15 µg/ml2 will be referred to systematic biopsy despite a negative MRI (Distler 2017). A random sample of those men with PSA≥3 but (1) negative kallikrein panel and/or (2) negative MRI will also undergo systematic biopsy for assessment of diagnostic accuracy. A small number of men will not be eligible for MRI due to metal implants, and for them transrectal ultrasound (TRUS) will be used instead.

The prostate biopsies are evaluated by trained, experienced uropathologists under supervision of Tuomas Mirtti in Helsinki and Paula Kujala in Tampere. Structured and standardised reports are used, with each core is evaluated separately and exact location recorded. Training and standardisation will be conducted

using virtual microscopy methods developed in Tampere (prof. Jorma Isola). The original pathology reports are retrieved from the system and incorporated into the trial database in an automated fashion.

The expected numbers of screen-detected cases in the first round is 290 (55% of men referred to MRI, or 2.6% of screening participants). Of the screen-detected cases, 75-80% are expected to be Gleason≥7. Treatment of the detected cases is centralised at the Helsinki and Tampere university hospitals and follows the Finnish guidelines. Primary management options for localised cancer currently include active surveillance (according to EAU-ESTRO-SIOG guideline), radical prostatectomy (robot-assisted), external radiotherapy or brachytherapy. For locally advanced cases, combined modality with radiotherapy and endocrine treatment is treatment of choice. For metastatic cases (and localised cases with contraindications curative treatment), endocrine treatment (LHRH antagonists, antiandrogens) and watchful waiting are used. Any future developments in treatment policy is implemented in management of the cases in both arms of the trial. This is an ethical imperative and also satisfies the scientific requirement of similar treatment, given disease and patient characteristics, in both trial arms in order to avoid bias due to treatment.

Data collection

Cost data will also be collected for economic analyses, including procedure codes, DRG group and actual billing information for each hospital visit (including outpatient visits) from the hospital database. Questionnaire surveys will be used for random samples of men in both arms to obtain utility estimates (with the 15D instrument) and information on out of pocket health care costs. For indirect costs, data on sickness benefits will be obtained from the Social Insurance Institution (SII) database (the governmental agency providing all benefits through the national comprehensive health insurance system).

Adverse effects of biopsy are evaluated, with questionnaire surveys of complications, including assessment of bleeding, lower urinary tract symptoms (LUTS) and erectile dysfunction (ED) at four weeks (with baseline data on LUTS and ED prior to biopsy collected at the time of the examination), as well as hospitalisation data on septic infections. The questionnaire instruments used are ICIQ-MLUTS for LUTS and IIEF-5 for ED; in addition, data on antibiotic prescriptions following biopsy will be obtained from the SII nationwide prescription database.

Information on prostate cancer cases diagnosed in the entire study population will be obtained from the Finnish Cancer Registry (with>99% coverage of cancer cases in Finland), so that interval cancers as well those in non-participants and the control arm can be identified and intention to screen analyses of prostate cancer incidence in the entire study population conducted. Detailed clinical information on Gleason score, PSA at diagnosis, treatment and other features will be retrieved from the electronic medical record using the ECRF software customised for the trial.

As for patient-oriented outcomes, assessment of the generic quality of life (SF12 and 15D) will be conducted for a sample of men at each stage of the screening process. The validated Extended Prostate Cancer Index Composite (EPIC) questionnaire is used for monitoring quality of life and adverse effects of treatment in all men within trial diagnosed with prostate cancer (regardless of arm, hospital and treatment).

Follow-up for vital status and emigration is through Population Registry, and causes of death will be obtained from Statistics Finland. We have previously shown excellent agreement between the official causes of death and a blinded cause of death adjudication panel (kappa 0.9, Mäkinen 2008, Kilpeläinen 2016) and no adjudication committee will be used, because of the demonstrated validity of the Statistics Finland cause of death data.

5. TRIAL SCHEDULE

The study will commence in 2018, with randomisation of half of the men during the first year. The first screening round will be completed in 2019. A summary of the timeline is shown in Table 1.

Table 1. Planned schedule of the trial conduct

	Trial years 1-2	Years 3-4	Years 5-6	Years 7-8	Years 9-10
All subjects	Recruitment				
All participants	1st screening				
Men with PSA>3		2nd screen	3rd screen	4th screen	5th screen
Men with PSA 1.5-3			2nd screen		3rd screen
Men with PSA<1.5				2nd screen	

6. POWER CALCULATIONS & STATISTICAL ANALYSIS PLAN

The main outcome of the trial is prostate cancer mortality. An intention to screen analysis will be performed, with all men in the groups defined by random allocation, regardless of compliance. This is the gold standard for randomised trials, as it gives an unbiased effectiveness estimate, which is realistic for application (though lower and 'diluted' compared with an efficacy estimate based on compliers only). The only post-randomisation exclusions will be based on prostate cancer diagnosis prior to entry, as inevitably some recent cases will be found out only after randomisation. Follow-up starts at randomisation, and ends at death. Dates of death are registered also for men who have emigrated, but in case of a missing cause of death, the event will be censored. Cox regression will be used with prostate cancer death as the outcome. Randomisation is expected to result in balanced distribution of age, with no need for adjustment.

Based on the most recent age-specific population rates (2010-2014) and the age structure of the male population, the expected number of prostate cancer deaths in the control arm is 222 at 10 years, with sufficient statistical power ($1-\beta=0.8$, $\alpha=0.05$ two-sided) to detect a 33% reduction (hazard ratio, $HR\leq0.67$ for the screening arm, cumulative mortality 29 vs. 44 per 10,000 men) in prostate cancer mortality. This is calculated considering the overall mortality from other causes (ranging from 7 per 100,000 at age 55 to 210 per 100,000 at age 75), yielding a total of 475,000 person-years of follow-up. At 15 years of follow-up, the expected number of deaths in the control arm is 451, with a minimal detectable effect size of 26% ($HR\leq0.74$, 67 vs. 90 deaths per 10,000 men, based 912,800 person-years of follow-up).

A secondary analysis of prostate cancer mortality will be performed using the Cuzick method to correct for non-compliance and contamination (Cuzick 1997, Roobol 2009, Kilpeläinen 2015).

Intermediate outcomes include cumulative incidence of advanced (T3-T4 or M1) prostate cancer (number of cases relative to population size, not using incidence density to avoid the lead-time bias due to early

detection by screening), and cumulative incidence of low-risk cancer (Gleason<7) as an indicator of overdiagnosis.

Analysis of screening test performance, including sensitivity and specificity, positive and negative predictive value will be conducted, assessing also the accuracy of 4Kscore and MRI. Interval cancer incidence will be analysed as an indicator of program sensitivity, besides participation and its determinants. Cancer detection and screening test performance will also be evaluated by major risk factors including family history, genetic risk and major lifestyle determinants.

Short-term quality of life impact of screening will be evaluated among screening participants, and adverse effects of prostate biopsy analysed as outlined above. Quality of life and psychosocial impacts among men diagnosed with prostate will be assessed using the EPIC instrument and the Memorial Anxiety Scale for Prostate Cancer (MAX-PC), enrolling screen-detected and interval cases, as well as those diagnosed among non-participants and in the control arm. Follow-up will be carried out at diagnosis and 6, 12 and 48 months after it. A mobile phone app is used to collect data on patient-oriented outcomes.

Economic evaluation will commence with cost analysis, and the final analysis of incremental cost effectiveness ratio (with a decision analysis) can be conducted only once data on both long-term cost and real outcome data on both utilities (quality-adjusted life-years) and mortality are available. However, early assessment can give some predictions about potential effects size required for the intervention to be regarded as cost-effective.

7. ETHICAL ISSUES

Screening is by definition targeting people free of the target condition, and the benefit per participant is nearly always small. Consequently, any risk must also be small to justify the intervention. The primary screening tests are based on biochemical analyses of a blood sample, with minimal risks. All the other procedures in the trial (diagnosis and treatment among screen-positive men) are conducted in accordance with the normal standards of care and pose therefore no ethical issues. The study protocol will be submitted for review by the ethical committees of the Uusimaa and Pirkanmaa hospital districts. Permission for cancer registry linkage will be obtained from the National Institute for Health and Welfare (THL). Written informed consent is obtained from all screening participants for both the intervention and collection of data from electronic health records and health care registers for the analyses of the trial (cancer registry, cause of death registry, hospital discharge registry, regional laboratory and national prescription databases). Information about screening covers probabilities of abnormal screening test result, prostate cancer detection at biopsy given screen-positive result, frequency of biopsy complications, treatment options with their adverse effects, overdiagnosis in screening, as well as mortality impact of screening. The information is provided as a decision aid with graphically presented numerical results, modified from the earlier Finnish screening trial. A trial website will also be constructed to provide information on the progress of the study for the participants. According to the Finnish regulation, no consent is required for non-compliers or men in the control arm, but register data can be collected with permission from THL after an ethics committee review.

The trial can be stopped early for benefit or harm, if evidence for difference between the arms is obtained. For assessment of harm from the intervention, biopsy and treatment complications will be analysed, as well as incidence of advanced cancer (as an indicator of cancers missed by screening) and Gleason <7 cancer (as indicator of overdiagnosis) by arm are performed annually for the data monitoring committee. Monitoring of the main end-point (prostate cancer mortality) is carried out by an external committee. Interim analyses

are performed at five and eight years in accordance with sequential analysis rules (total alpha spending 0.005 of the two interim analyses according O'Brien-Fleming).

The study question satisfies equipoise criteria, i.e. balance of uncertainty (superiority of the evaluated intervention versus standard practice is not established). All procedures applied represent viable clinical options meeting current standards of care. The two screening tests based on a blood sample constitute the only experimental procedure performed solely for research purposes. Hence, the allocation based on randomisation is the only deviation from the normal health care and patient management. Screen-positive men are referred to diagnostic examinations just as any other patients, and undergo the procedures similar to other patients. This means that the results are readily applicable to public health policy. The generalisability is further enhanced by the Zelen-type randomisation, with consent sought only from the men in the intervention arm after randomisation. This ensures ethical conduct of research, as nobody is denied access to proper health care by the randomisation.

8. RISKS

The fundamental idea of the trial is the uncertainty about the effects of prostate cancer screening and the rationale rests on the assumption that the approaches, which have never been evaluated in a randomised trial and never combined into a single protocol, can be successfully integrated into a winning formula. However, a RCT has the advantage that even in a negative result is informative, i.e. proving that the proposed concept does not work, advances our understanding of the issue, here prostate cancer screening. The study is unique, and the closest counterparts either use PSA alone and lack one layer of risk stratification (Göteborg2 study) or have entirely failed to use randomisation (Sthlm3 study) and hence fall short of the scope of the present proposal.

Among the less fundamental issues, we expect a similar participation proportion as in the previous Finnish screening trial (67% at the initial round, and a similar level retained in the subsequent rounds).

The anticipated effect size, i.e. magnitude of mortality reduction achievable may be overstated. In the European screening trial (ERSPC), the relative reduction in prostate cancer has remained at 20% from 9 years to 13 years of follow-up (Schröder 2009, 2014). If the two risk stratification methods used for screening (kallikrein panel and MRI) both have 90% sensitivity for clinically relevant prostate cancer, a realistic effect size might be 0.9*0.9*0.2=0.16 i.e. 16% reduction. However, the shorter screening interval for men with PSA>3 is likely to compensate this at least partly. For instance, the majority of prostate cancer deaths up to 19 years of follow-up was among men with a baseline PSA >3 in the screening participants in the previous Finnish trial, suggesting that the shorter screening interval in such men is likely to enhance the screening effect.

Prostate cancer mortality has declined in Finland from 60 to 50 per 100,000 in 2002-2014, and the power calculations are based on the most recent mortality rates. If the decreasing trends continue (with further advances in treatment outcomes), the observed death rates could be lower by up to a quarter relative to those used in the power calculations. Assuming 469 prostate cancer deaths from prostate cancer at 15 years, instead of 581, would reduce the statistical power and allow us to detect a minimal difference of 31% (HR≤0.69).

Active surveillance, with curative treatment postponed until signs of disease progression, has gained popularity in management of prostate cancer recently (Moschini 2017). Though treatment results of active surveillance have been highly favourable, it has a potential to decrease the screening impact as in some cases it may results in undertreatment and poor outcomes.

Statistical power can also be reduced by dilution of the contrast in the intervention between the arms. Spill-over means intervention becoming available also to the control arm during the trial. PSA testing prior to baseline would also reduce the expected impact of screening. PSA testing frequency (including diagnostic assessment of men with urological complaints, as well as monitoring of men treated for prostate cancer, besides testing for screening purposes) was very high (up to >20%) in Finland after 2000 (Kilpeläinen 2017), but the stark reduction in prostate cancer incidence by a quarter in the past 10 years clearly suggest a decline its popularity.

Any future changes on clinical practice regarding diagnosis or management of prostate cancer will be incorporated in the protocol and applied to all men in the study. For instance, a faster, simpler MRI protocol (biparametric imaging) is being developed, and recent evidence indicates that adding abiraterone in the first-line endocrine treatment of metastatic prostate cancer may improve treatment results.

9. FUNDING

The ProScreen trial has received preliminary funding from the Academy of Finland, Pirkanmaa Cancer Society, and Finnish Cancer organizations.

10. REFERENCES

Ahmed H, Bosaily AE, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS). Lancet 2017;389:815-22

Al-Olama AA, Kote-Jarai Z, Bernd SI, et al. A meta-analysis of 87.040 individuals identified 23 new susceptibility loci for prostate cancer. Nat Genet 2014;46:1103-9

Alberts AR, Schoots IG, Bokhorst LP, et al. Risk-based patient selection for MRI-targeted prostate biopsy after negative TRUS-guided biopsy avoids unnecessary MRI scans. Eur Urol 2016;69:1129-34

Auvinen A, Hakama M. Cancer screening: Theory and practice. International Encyclopedia of Public Health, Elsevier 2016

Auvinen A, Moss SM, Tammela TL, et al. Absolute effect of prostate cancer screening. Clin Cancer Res 2016;22:243-249

Barocas DA, Alvarez J, Resnick MJ, et al. Association between radiation therapy, surgery, or observation for localized prostate cancer and patient-reported outcomes after 3 years. JAMA 2017;317:1126-40

Beckmann KR, Lynch JW, Hiller JE, et al. A novel case-control design to estimate the extent of over-diagnosis of breast cancer due to organised population-based mammography screening. Int J Cancer 2015;136:1411-21

Bell KJ, del Mar C, Wright G, et al. Prevalence of incidental prostate cancer: A systematic review of autopsy studies. Int J Cancer 2015;137:1749-57

Beral V, Alexander M, Duffy S, et al. The number of women who would need to be screened regularly to prevent one death from breast cancer. J Med Screen 2011;18:210-2

Bjurlin MA, Carter HB, Schellhammer P, et al. Optimisation of initial prostate biopsy in clinical practice. J Urol 2013;189:2039-43

Braun K, Sjöberg DD, Vickers AJ, et al. A four-kallikrein panel predicts high-grade cancer on biopsy. Eur Urol 2016;69:505-11

Bryant RJ, Sjöberg DD, Vickers AJ, et al. Predicting high-grade cancer at ten-core biopsy using a four-kallikrein panel markers. J Natl Cancer Inst 2015;107:djv095

Carlsson S, Maschino M, Schröder FH, et al. Predictive value of four kallikrein markers for pathologically insignificant compared with aggressive prostate cancer. Eur Urol 2013;64:693-9

Carter HB, Partin AW, Walsh PC, et al. Gleason 6 adenocarcinoma: should it be labelled as cancer? J Clin Oncol 2012;30:4264-6

Chen RC, Basak R, Meyer AM; et al. Association between choice of radical prostatectomy, external beam radiotherapy, brachytherapy or active surveillance and patient-reported quality of life among men with localized prostate cancer. JAMA 2017;317:1141-50

Cuzick J, Edwards R, Segnan N. Adjusting for non-compliance and contamination in randomised clinical trials. Stat Med 1997;16:1017-291997

Distler FA, Radtke JP, Bonekamp D, et al. The value of PSA density in combination with PIRADS for the accuracy of prostate cancer prediction. J Urol 2017 (E-publication ahead of print)

Donovan J, Hamdy FC, Lane JA, et al. Patient-reported outcomes after monitoring, surgery, or radiotherapy for prostate cancer. N Engl J Med 2016;375:1425-37

Draisma G, Etzioni R, Tsodikov A, et al. Lead time and overdiagnosis in prostate-specific antigen screening. J Natl Cancer Inst 2009;101:374-83

Eeles R, Olama AA, Benlloch S, et al. Identification of 23 new susceptibility loci using the iCOGS custom genotyping array. Nat Genet 2013;45:385-91

Epstein JI, Zelefsky MJ, Sjöberg DD et al. A contemporary prostate cancer grade grouping system. Eur Urol 2016;69:428-4

Esserman LJ, Thomas IM, Reid B, et al. Addressing overdiagnosis and overtreatment of cancer. Lancet Oncol 2014;15:e234-42

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

Falk RS, Hofvind S, Skaane P, Haldorsen T. Overdiagnosis of invasive breast cancer due to mammography screening. Int J Cancer 2013;133:705-12

Fütterer JJ, Briganti A, de Visschere P, et al. Can clinically significant prostate cancer be detected with multiparametric magnetic resonance imaging? A systematic review of the literature. Eur Urol 2016;68:1045-53

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

GloboCan 2012: Ferlay J, Soerjomataram I, Ervik M, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available from: http://globocan.iarc.fr

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

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

Heijnsdijk EA, de Carvalho TM, Auvinen A, et al. Cost-effectiveness of prostate cancer screening. J Natl Cancer Inst 2014;107:366

Jahn JL, Giovannucci, EL, Stampfer MJ. The high prevalence of undiagnosed prostate cancer at autopsy. Int J Cancer 2015;137:2795-802

Kekki H, Peltola M, van Vliet S, et al. Improved specificity in PSA assay using Aleuria aurantia lectin coated Eu-nanoparticles for detection. Clin Biochem 2017;50:54-61

Kilpeläinen T, Tammela TL, Malila N et al. Screening failures in FinRSPC trial. Int J Cancer 2015;146:2437-43

Kilpeläinen TP, Mäkinen T, Karhunen PJ, et al. Estimating bias in causes of death in the Finnish Randomised Study of Screening for Prostate Cancer. Cancer Epidemiol 2016;45:1-5

Kilpeläinen TP, Pogodin-Hannolainen D, Kemppainen K, et al. Estimate of opportunistic prostate specific antigen testing in the Finnish Randomised Study of Screening for Prostate Cancer. J Urol 2017;198:50-57

Lao C, Brown C, Rouse P, et al. Economic evaluation of prostate cancer screening: a systematic review. Future Oncol 2015;11:467-77

Leapman MS, Cowan JE, Simo J, et al. Application of a prognostic Gleaosn grade grouping system to assess distant prostate cancer outcomes. Eur Urol 2017;71:750-9

Lin JS, Piper MA, Perdue LA, et al. Screening for colorectal cancer: Updated evidence report and systematic review for the US PSTF. JAMA 2016;315:2576-94

Loeb S, Folkvaljon Y, Robinson D, et al. Evaluation of the 2015 Gleason Grade Groups in a nation-wide population-based cohort. Eur Urol 2016;69:1135-41

Mäkinen T, Karhunen P, Aro J, et al. Assessment of causes of death in prostate cancer screening trial. Int J Cancer 2008;122:413-7

McPhail S, Johnson S, Greenberg D, et al. Stage at diagnosis and early mortality from cancer in England. Br J Cancer 2015;112:S108-15

Moschini M, Carroll PR, Eggener SE, et al. Low-risk prostate cancer: Identification, management and outcomes. Eur Urol 2017 (E-publication ahead of print)

Nevalainen J, Stenman UH, Tammela TL, et al. What explains the differences between centres in the European screening trial? Cancer Epidemiol 2017;46:14-19

Palma A, Lounsbury DW, Schlecht NF, Agalliu I. A system dynamics model of serum prostate-specific antigen screening for prostate cancer. Am J Epidemiol 2016;183:227-36

Parekh DJ, Punne S, Sjöberg DD, et al. A multi-institutional prospective trial in the USA confirms that the 4Kscore accurately identifies men with high-grade prostate cancer. Eur Urol 2015;68:464-70

Pashayan N, Pharoah PD, Schleutker J, Reducing overdiagnosis by polygenic risk-stratified screening. Br J Cancer 2015;113:1086-93

Pinsky PF, Prorok PC, Yu K, et al. Extended mortality results for prostate cancer screening in the PLCO trial with median follow-up of 15 years. Cancer 2017;123:592-9

Pierorazio PM, Mulins JK, Eifler JB, et al. Prognostic Gleason grade grouping: Data based on the modified Gleaosn scoring system. BJU Int 2013;112:751-7

Randazzo M, Müller A, Carlsson S, et al. A positive family history as a risk factor for prostate cancer in a population-based study with organized PSA screening. BJU Int 2016;117:576-83

Roobol M, Kerkhof M, Schröder FH, et al. Prostate cancer mortality reduction by PSA-based screening adjusted for non-attendance and contamination in the European randomized Study of Screening for Prostate Cancer. Eur Urol 2009;56:584-91

Roine A, Veskimäe E, Tuokko A, et al. Detection of prostate cancer by an electronic nose: A proof of principle study. J Urol 2014;192:230-4

Saarimäki L, Tammela TL, Määttänen L, et al. Family history in the Finnish prostate cancer screening trial. Int J Cancer 2015;136:2172-7

Schoots IG, Roobol MJ, Nieboer D, et al. Magnetic resonance imaging-targeted biopsy may enhance the accuracy of significant prostate cancer detection: Systematic review. Eur Urol 2015;68:438-50

Schröder FH, Hugosson J, Roobol M, et al. Screening and prostate cancer mortality in a randomized European study. N Engl J Med 2009;360:1320-28

Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate cancer mortality: Results of the European trial at 13 years of follow-up. Lancet 2014;384:2027-35

Seikkula HA, Kaipia AJ, Rantanen ME, et al. Stage-specific mortality and survival trends in prostate cancer patients in Finland before and after introduction of PSA. Acta Oncol 2017;56:971-7

Szulkin R, Whitington T, Eklund M, et al. Prediction of individual genetic risk to prostate cancer using a polygenic score. Prostate 2015;75:1467-74

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

Welch HG, Prorok PC, O'Malley AJ, Kramer BS Breast cancer tumor size, overdiagnosis and mammography screening effectiveness. N Engl J Med 2016;375:1438-47

Wu GH, Auvinen A, Määttänen L, et al. Number of screens for overdetection as an indicator of absolute risk of overdetection in cancer screening. Int J Cancer 2012;131:1367-76