Prostate Cancer Screening

-Aspects of Overdiagnosis

Rebecka Arnsrud Godtman

Department of Urology
Institute of Clinical Sciences
Sahlgrenska Academy at University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2014

Prostate Cancer Screening

r.godtman@gmail.com

ISBN 978-91-628-9224-1, ISBN 978-91-628-9251-7
http://hdl.handle.net/2077/36913
Printed in Gothenburg, Sweden 2014
Ale tryckteam

To my family

Prostate Cancer Screening

-Aspects of Overdiagnosis

Rebecka Arnsrud Godtman

Department of Urology, Institute of Clinical Sciences
Sahlgrenska Academy at University of Gothenburg
Göteborg, Sweden

ABSTRACT

The overall aim of this thesis is to explore aspects of overdiagnosis, i.e. the diagnosis of a tumor that in the absence of screening would never have been diagnosed, in prostate cancer (PC) screening. The four papers in this thesis all emerge from the Göteborg randomized population-based PC screening trial, in which 10,000 men were invited to biennial prostate-specific antigen (PSA)-screening between 1995 and 2014 and 10,000 non-invited constituted a control group. In paper I, the accuracy of cause of death (COD) certificates, for men with PC, is evaluated by comparison with the COD as assigned by an independent committee after blinded review of medical records. Paper II assesses outcomes for men with screen-detected PC managed with, so called "active surveillance". In paper III, organized screening is compared with opportunistic screening with respect to effectiveness in reducing PC mortality, measured as the number needed to invite (NNI) to screening and overdiagnosis, measured as number needed to diagnose (NND) to prevent one man from dying from PC. Paper IV investigates the risk of being diagnosed with PC depending on age at screening and the number of screens. The overall agreement between COD certificates and the committee was 96%. A large proportion of men screen-detected PC has low-risk PC (60%) and could safely be managed with active surveillance, at least with intermediate follow-up. Organized screening was more effective in reducing PC mortality and was associated with less overdiagnosis than opportunistic screening (NNI 139, NND 13 versus NNI 493, NNI 23). The risk of being diagnosed with PC increased dramatically with age but there was no apparent relation to the number of screens. From this thesis it can be concluded that Swedish COD certificates have a high accuracy and can be used for COD determination for men with PC, at least in the age-range studied (50-64 years old at the start of screening). Active surveillance appears safe for men with low-risk PC and should be used as a treatment strategy in order to reduce overtreatment. In order to reduce overdiagnosis and improve the benefit harm ratio of PC screening, screening should be conducted within the frameworks of an organized program where "younger" men could be screened relatively intense but where "older" men are screened more selectively.

Keywords: active surveillance, age, cause of death, opportunistic, organized, overdiagnosis, prostate cancer, prostate-specific antigen, risk factors, screening, screening interval

ISBN: 978-91-628-9224-1, ISBN 978-91-628-9251-7 http://hdl.handle.net/2077/36913

SAMMANFATTNING PÅ SVENSKA

Bakgrund

Denna avhandling har som övergripande mål att studera olika aspekter av överdiagnostik vid screening för prostatacancer (PC) med blodprovet prostataspecifikt antigen (PSA). Med överdiagnostik menas diagnos av en cancer som i avsaknad av screening aldrig skulle ha gett symptom eller ha upptäckts. Överdiagnostik leder till att ”friska” män får en cancerdiagnos och riskerar att behandlas i onödan. Den botande behandlingen för PC (operation eller strålbehandling) är förknippad med många biverkningar som till exempel nedsatt potens, urinläckage och tarmbesvär. Överdiagnostik och dess konsekvenser är huvudanledningen till att allmän screening för PC inte har införts i Sverige trots att det finns starka belägg för att PSA-screening skulle kunna minska dödligheten i PC. Denna avhandling består av fyra delarbeten som alla är sprungna ur en screeningstudie för PC i Göteborg. Denna studie startades 1995 då 10,000 män, födda mellan 1930 och 1944, lottades till regelbundna PSA-kontroller och 10,000 män lottades till att utgöra en kontrollgrupp som inte inbjöds. Våren 2014 avslutades den 10:e och sista screeningomgången. Avhandlingens fyra delarbeten syftar till att besvara följande frågeställningar:

- Kan svenska dödsorsaksintyg användas för utvärdering av PC dödlighet i screeningstudien i Göteborg trots stor skillnad i överdiagnostik mellan armarna?
- Kan överbehandling minska genom aktiv övervakning? Är aktiv övervakning ett säkert behandlingsalternativ för utvalda män med screeningupptäckt PC?
- Skiljer sig organiserad och opportunistisk screening åt avseende effektivitet i att minska dödligheten i PC och risken för överdiagnostik?
- Hur påverkar ålder och antal gånger en man PSA-testas för risken att få diagnosen PC?

Metoder

I det första delarbetet insamlades journaler och dödsorsaksintyg för alla män med PC-diagnos som hade avlidit mellan 1995 och 2008. En expertkommitté bestående av tre erfarna urologer granskade sedan materialet och fastställde dödsorsaken med hjälp av en algoritm (flödesschema). Expertkommitténs utlåtande jämfördes därefter med dödsorsaken på dödorsaksintyget. I det andra delarbetet studeras de män med screeningupptäckt PC som inte omedelbart genomgick aktiv behandling

med operation eller strålbehandling, utan som följdes med så kallad aktiv monitorering mellan åren 1995 och 2010. Aktiv övervakning är en behandlingsstrategi som syftar till att minska onödig behandling av screeningupptäckt PC. Med denna strategi följs mannen med regelbundna kontroller och först om tumören visar tecken på att växa eller bli mer aggressiv går man vidare med operation eller strålbehandling. Förhoppningen är att mannen helt kan avstå alternativt skjuta upp, behandling ett antal år utan att chansen till bot missas. I det tredje delarbetet jämförs organiserad screening med opportunistisk (oorganiserad) screening avseende förmåga att minska dödligheten i PC och risken för överdiagnostik. Screeninggruppen i Göteborgsstudien har genomgått organiserad screening och kontrollgruppen har under samma period exponerats för opportunistisk screening, det vill säga PSA-testning på vårdcentralen, i samband med hälsokontroller eller som del i utredning av till exempel vattenkastningsbesvär. Genom att jämföra med historiska data från 1990-94 (innan PSA var utbrett som screeningtest) kunde vi studera hur organiserad och opportunistisk screening påverkat incidens (antal nya PC-fall över tid) och dödlighet i PC. I det fjärde arbetet studeras de män som deltagit i alla screeningomgångar de inbjudits till, vilket kunde variera mellan 3 och 10 screeningtillfällen, beroende på hur gamla de var vid studiestart. Eftersom männen hade genomgått olika antal PSA-test vid olika åldrar kunde vi jämföra hur stor risken var att bli diagnostiserad med PC, och därmed också risken att bli överdiagnostiserad, beroende på ålder och antalet gånger en man tagit PSA.

Resultat och kommentarer

I: Dödsorsaken angiven på dödsorsaksintygen stämde till 96% överens med den dödsorsak som kommittén hade angett. Då de fall där dödsorsaken på intyget och kommitténs beslut inte överensstämde var få kunde inte någon riskfaktor för ett felaktigt intyg fastställas. Resultaten visar att svenska dödsorsaksintyg för män med PC håller hög kvalitet, åtminstone inom ramen för den studerade åldersgruppen (50-64 år vid start av screeningen). När männen i studien blir äldre ökar annan sjuklighet vilket eventuellt kan försämra kvaliteten på dödsorsaksintygen.

II: En stor andel (60%) av de män som diagnostiserats med screeningupptäckt PC har cancer av lågrisktyp. Aktiv övervakning förefaller vara en säker monitoreringsstrategi för dessa män, i alla fall under en begränsad tid (i denna studie 6 år). För män med tumörer av en högre riskkategori tycks aktiv övervakning vara mer riskfyllt. Om en man med denna tumörtyp önskar aktiv övervakning bör han tydligt informeras om att det finns risk att missa chansen att bli botad vid senarelagd operation eller strålbehandling.

III: Organiserad screening var mer effektiv än opportunistisk screening när det gällde att minska dödligheten i PC. Med en uppföljning på 18 år behövde 139 män bjudas in till organiserad screening för att förhindra ett dödsfall i PC, medan motsvarande siffra för opportunistisk screening var att 493 män behövde exponerats för denna screeningform. Dessutom resulterade opportunistisk screening i mer överdiagnostik än organiserad, 23 män behövde diagnostiseras med PC för att förhindra ett dödsfall med medan denna siffra var 13 för organiserad screening.

IV: Risken att bli diagnostiserad med PC var kraftigt beroende av ålder medan antalet gånger en man hade tagit PSA-test var av mindre betydelse. Om en man till exempel hade kontrollerat PSA fem gånger vid 60 års ålder var risken för PC 8.4% medan motsvarande risk vid 65 års ålder var 13% och vid 70 år 21%. Resultaten indikerar att risken för överdiagnostik är mer kopplad till åldern för när en man slutar kontrollera PSA än hur många gånger en man kontrollerat sitt PSA.

Slutsatser

Svenska dödsorsaksintyg för män med PC håller hög kvalitet och kan användas som underlag för dödsorsaksbestämning i screeningstudier för PC. Överdiagnostik är vanligt vid PSA-screening och ökar kraftigt med stigande ålder. Om en välinformerad man önskar PSA-testning bör detta ske inom ramen för ett organiserat program med täta intervall och noggrann uppföljning. Aktiv övervakning bör vara ett alternativ för utvalda män med screeningupptäckt lågrisk PC i syftet att minska onödig behandling. Möjliga förbättringsområden för PSA-screening som skulle kunna förbättra balansen mellan fördelar och nackdelar är:

-organisera PSA-screeningen inom ramen för ett screeningprogram.

-screena mer selektivt; undvik screening av äldre män och de med annan sjuklighet

-undvik onödig omedelbar aktiv behandling med operation eller strålbehandling för män med cancer av lågrisktyp genom att erbjuda aktiv övervakning.

Framtiden

Det pågår mycket forskning för att hitta bättre verktyg för screening och tidig diagnostik av PC. Nya biomarkörer, genetiska test och bilddiagnostiska metoder så som multiparametrisk magnetkameraundersökning verkar lovande inför framtiden. Det ultimata screeningtestet/undersökningsmetoden bör vara

ickeinvasivt, billigt, ha en hög sensitivitet och specificitet för PC och undvika att diagnostisera cancer som aldrig skulle gett symptom i frånvaro av screening.

LIST OF PAPERS

This thesis is based on the following studies, which are referred to in the text by their Roman numerals.

I. **Godtman R**, Holmberg E, Stranne J, Hugosson J. High accuracy of Swedish death certificates in men participating in screening for prostate cancer: a comparative study of official death certificates with a cause of death committee using a standardized algorithm. *Scand J Urol Nephrol.* 2011;45(4):226-32.

II. **Arnsrud Godtman R**, Holmberg E, Khatami A, Stranne J, Hugosson J. Outcome following active surveillance of men with screen-detected prostate cancer. Results from the Göteborg randomized population-based prostate cancer screening trial" *Eur Urol*. 2013;63(1):101-7.

III. **Arnsrud Godtman R**, Holmberg E, Lilja H, Stranne J, Hugosson J. Opportunistic testing versus organized prostate-specific antigen screening, outcome after 18 years in the Göteborg Randomised Population-Based Prostate Cancer Screening Trial. *Submitted.*

IV. **Anrsrud Godtman R**, Carlsson S, Holmberg E, Stranne J, Hugosson J. Age at termination of screening – the most important risk factor for (over) diagnosis in screening for prostate cancer. Results from the Göteborg Randomised Population-based Prostate Cancer Screening Trial.In manuscript.

CONTENT

ABBREVIATIONS

AUA	American Urological Association
AUC	Area under the curve
BPH	Benign prostatic hyperplasia
CAPRA	Cancer of the Prostate Risk Assessment
CaPSURE	Cancer of the Prostate Strategic Urologic Research Endeavor
CT	Computed tomography
COD	Cause of death
DRE	Digital rectal examination
EAU	European Association of Urology
EBRT	External beam radiotherapy
EORTC	European Organization for Research and Treatment of Cancer
HRQoL	Health-related quality-of-life
HDR	High-dose rate
IARC	International Agency of Research of Cancer
ISUP	International Society of Urological Pathology
LDR	Low-dose rate
LUTS	Lower urinary tract symptoms
MRI	Magnetic resonance imaging
mp-MRI	Multiparametric magnetic resonance imaging

NND	Number needed to diagnose to prevent one prostate cancer death
NNI	Number needed to invite to screening to prevent one prostate cancer death
NPV	Negative predictive value
NSO	Number of screens for overdetection
PC	Prostate cancer
PCBaSe	Prostate Cancer Data Base
PCPT	Prostate Cancer Prevention Trial
PET-CT	Positron emission tomography-computed tomography
ProtecT	Prostate testing for cancer and Treatment
PRIAS	Prostate Cancer Research International Active Surveillance
PSA	Prostate-specific antigen
PSAD	Prostate-specific antigen density
PSADT	Prostate-specific antigen doubling time
PLCO	Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial
PPV	Positive predictive value
QALY	Quality adjusted life-years
QoL	Quality-of-life
RCC	Regional Cancer Center
RCT	Randomized controlled trial

RTOG	Radiation Therapy Oncology Group
SCR	Swedish Cancer Register
SE	Standard Error
SEER	Surveillance, Epidemiology, and End Result Program
SPR	Swedish Population Register
TRUS	Transrectal ultrasound
TUR-P	Trans-urethral resection of the prostate
UCSF	University of California San Francisco
USPSTF	US Preventive Services Task Force

1 INTRODUCTION

Modern medicine has strived towards detecting and treating conditions at earlier stages. With more sensitive tests and imaging techniques small tumors are now being detected, which in the absence of such examinations would never have been diagnosed during the lifetime of the host. This is referred to as *overdiagnosis*. Screening has been – and still is – an important strategy for early detection, as it enables the detection of a disease at an asymptomatic stage. However, during recent years, there has been a growing awareness that finding "everything" is not always desirable. The concept of overdiagnosis, and the associated concept of overtreatment, has gained attentiveness among medical professionals. However, much work remains and overdiagnosis is still not a term in Dorland's Medical Dictionary.[1]

This strive towards early detection has also influenced the field of urology and prostate cancer (hereafter referred to as PC). From being a highly lethal disease where most men were beyond the chance of cure by the time of diagnosis, the advent of prostate-specific antigen (PSA) as a screening test for PC has completely changed the clinical landscape of PC. Today, most men are diagnosed with early stage PC. With early diagnosis and treatment much suffering from advanced PC and many PC deaths can be prevented. However, similar to other forms of early detection strategies, PC screening is a double-edged sword; there are both pros and cons. A considerable proportion of those diagnosed with screen-detected PC have little to gain from being diagnosed or treated, because of the slow growing nature of certain PCs and/or from the risks of competing causes of death in older men. Whether or not organized screening for PC should be introduced in Sweden is an ongoing controversy. The main obstacles for implementing population-based screening are the high levels of overdiagnosis and overtreatment with current screening strategies. This difficult balance of benefits and harms is the rationale behind this thesis, which aims at exploring different aspects of overdiagnosis in screening for PC with PSA.

The four papers that constitute this thesis are all based on the Göteborg randomized population-based prostate cancer screening study.[2] This study started already in 1995 and at the time of writing this thesis the 10th and final screening round has just been completed. The Göteborg screening study is unique among screening studies for several reasons, mainly because it has a long follow up (20 years today). Another factor making the Göteborg study unique is the fact that when the study was launched in the mid 1990's, the Swedish male population constituted a previously unscreened population and

very few men had had a PSA-test. Therefore, the design of the Göteborg study will never be possible to replicate today, since PSA-testing is now more or less widespread. Thus, the Göteborg study constitutes an exclusive source of information regarding the effects of introducing organized screening on a previously unscreened population.

1.1 The prostate gland

The prostate is a small gland, normally the size of a walnut, located approximately 2 centimeters posterior to the pubic bone right below to the bladder. It is shaped like a truncated cone, enclosed by a capsule, with an anterior, posterior and lateral surface, a narrowed apex inferiorly and a broad base superiorly. The urethra runs through the prostate and the apex of the prostate is continuous with the urethral sphincter. Neurovascular bundles containing nerves controlling erectile function (potency) runs postero-lateral to the prostate in the lateral prostatic fascia making them vulnerable for being damaged when the prostate is removed surgically or treated with radiotherapy. The vas deferens and the two seminal vesicles are found posterior to the prostate and a small space, Denonvilliers fascia, separates them and the prostate from the rectal wall. The close contact with the rectal wall makes the prostate accessible for digital palpation and transrectal biopsies. A shallow groove palpable on rectal examination divides the gland in a right and left lobe.[3]

The prostate is composed of glandular elements and a fibromuscular stroma. Histologically it can be divided into three different zones (Figure 1); the transition zone from which benign prostatic hyperplasia (BPH) arises and where approximately 20% of all PC originate, the central zone where only 1-5% of all PC originates and the peripheral zone where the majority of the glandular tissue in located. This peripheral zone is also the zone where 70% of the PC arise and the zone commonly affected by chronic prostatitis.[3] The prostate glands consist of a single layer of secretory epithelial cells surrounded by a single layer of basal cells and a basal membrane.[4] The prostate produces 60% of the ejaculate and the prostatic secretion is believed to be important for the motility of the spermatozoa but the overall function of the prostate is principally unknown.

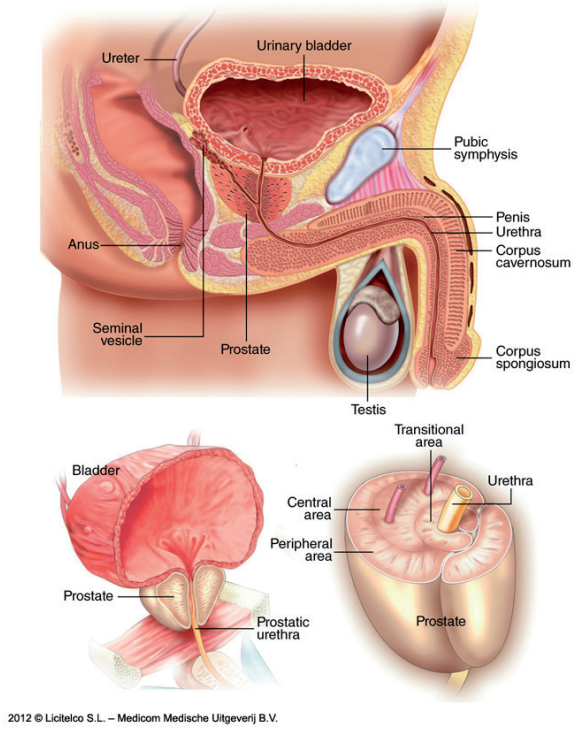

Figure 1. The prostate gland. (Reprinted with permission from AstraZeneca Oncology).

1.2 Prostate cancer

Prostate cancer is the most common cancer form (excluding non-melanoma skin cancer) in Swedish men. Every year, approximately 9000 men are diagnosed and PC is a major health concern. The age-adjusted PC mortality rate in Sweden is among the highest in the world. Approximately 2400 Swedish men die from PC every year.[5] What causes PC is largely unknown but older age, ethnicity and heredity are well known risk factors.[6] Prostate cancer incidence increases strongly with age, and the disease is uncommon before the age of 50 years.

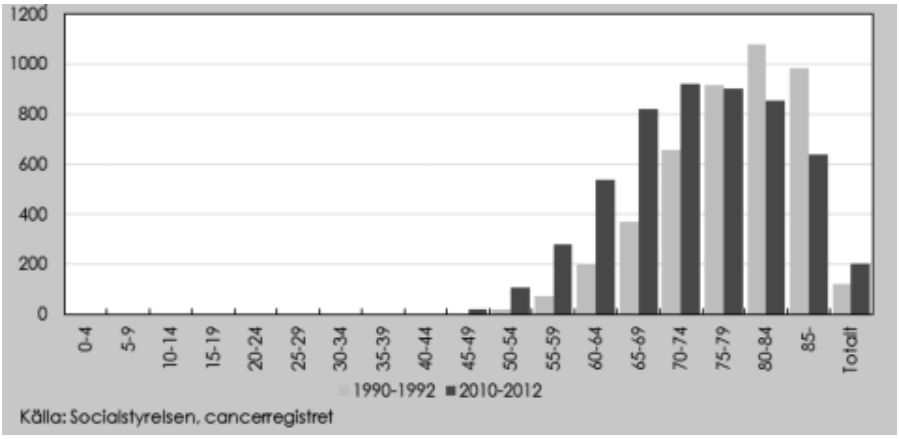

Figure 2. Prostate cancer, age-specific incidence, 1990-1992 and 2010-2012. Cases per 100 000, 3-year mean value. Adapted from: Cancerincidence i Sverige 2012,the National Board of Health and Welfare[5]

The median age at diagnosis in Sweden, as in many other countries with wide-spread PSA testing, has decreased from 74 years in 1995 to 69 years in 2005.[7] The age-span 65-69 years contains the greatest number of new cases (Figure 2).[5] Autopsy studies have confirmed the strong association between age and PC, showing that PC can be detected as early as in the 3rd decade of life. The prevalence of autopsy-detected PCs increases steadily with age, reaching 70-80% for men in their 80s.[8, 9] There are large geographic variations in both the incidence and mortality of PC. As with many other cancer forms heredity and environmental factors interact. A Western lifestyle with obesity, a high intake of dietary fat and red meat has been identified as a risk factor for developing PC, whereas a high intake of phyto-oestrogens and antioxidants have been suggested to have a protective effect. Chronic prostate inflammation have also been suggested to have a possible role in the development of PC.[6] Exogenous factors most certainly play an important role but the evidence available today is too weak and inconclusive to recommend any primary preventive measures. As previously mentioned, heredity is a very important factor and a large study based on the Swedish Family-Cancer Database showed that if the father had PC the risk for his son to be diagnosed with PC was 2-fold increased, but if three brothers were affected the risk was almost 18-fold increased.[10] True heredity PC, defined as three of more relatives with PC, or at least 2 close relatives who have developed early onset disease, is however, uncommon (approximately 9%). Men with hereditary PC usually have disease onset approximately six years earlier than spontaneous cases.[11]

1.2.1 Incidence and mortality trends

Prostate cancer is the second most common cancer in men and the sixth leading cause of death (COD) in men worldwide but there are large variations in incidence and mortality rates and trends over time.[12, 13] Incidence rates are highest in the high resource parts of the world such as North America and north-western Europe. On the contrary, mortality rates are among the highest in low-and medium resource countries such as Trinidad and Tobago and Cuba. Scandinavian countries also have high mortality rates. While the incidence rate is still increasing in most countries it has started to stabilize and decline in those countries which were among the first to adopt a widespread use of PSA (e.g. US and Canada). The greatest reductions in mortality rates are seen in high resource countries, while mortality is increasing in several countries in east and central Europe and South America.[13] Sweden is no exception to other high resource countries and in Sweden PC constitutes 32% of all cancer diagnosed.[14] Prostate cancer incidence was slowly increasing in Sweden until the mid to late 1990s. Thereafter the incidence rose dramatically and peaked in 2004 at a level of 223 new cases per 100 000 men (age-standardized).[15] The incidence now appears to have stabilized and even started to decline. Yet one in five Swedish men will receive a PC diagnosis during their lifetime.[15]

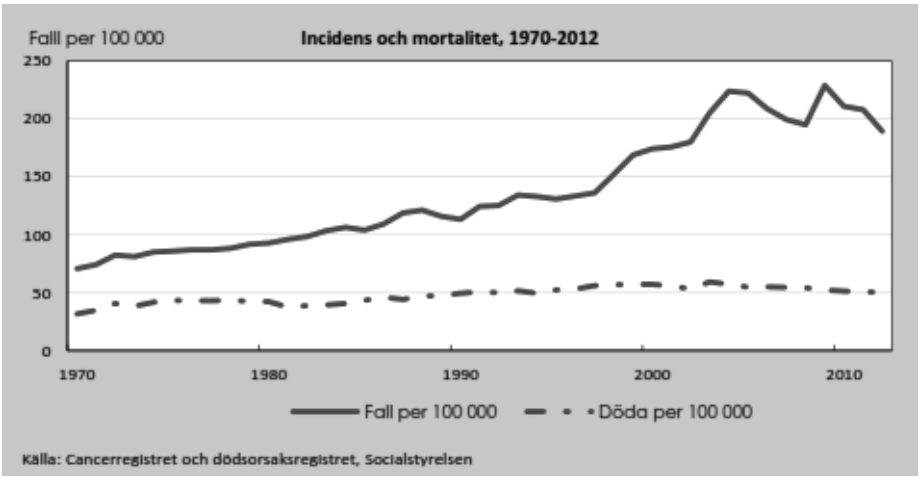

Figure 3. Prostate cancer incidence and mortality in Sweden 1970-2012. Number of prostate cancer cases and number of prostate cancer deaths per 100 000 (continuous line= incidence, dotted line=mortality). (Adapted from [5])

Prostate cancer mortality has not exhibited the same fluctuation as the incidence trend but has remained relatively stable since the 1960s. However, a small annual decrease of 2.2% in the age-standardized PC mortality rate has been observed during the last decade. The life time risk for PC death for Swedish men is 5-6%.[15]

Several factors contribute to the high PC incidence. An ageing population, increased awareness of prostate-related symptoms, better access to health care, increased usage of transurethral resection of the prostate (TUR-P) for BPH, an increase in the number of biopsy cores taken and a "true" incidence increase due to background risks such as exposure to dietary or environmental carcinogens are also contributing. However, most importantly, there is a direct relationship between the uptake of PSA use and PC incidence. Almost the entire incidence increase during the last 15-20 years can be explained by the detection of non-palpable (clinical stage "T1c") tumors in parallel with decreased number of men diagnosed with metastasized disease. As an example, Sweden's nationwide National Prostate Cancer Registry (NPCR) the proportion of men with low-risk tumors increase from 14% in year 1998 to 28% in 2012 while the proportion of men with distant metastases at diagnosis decreased from 25% to 13% during the same time period.[7] Another indication of earlier diagnosis is that the PSA-level at diagnosis has decreased from 23 ng/mL in 1998 to 9 ng/mL in 2012.[7] The proportion of men diagnosed with PC after a routine health check-up has increased from 29% in 2004 to 46% in 2012.[7] Future incidence trends are difficult to foresee as they are depending on future screening policies.

The decreasing PC mortality trend in the western world also has several possible explanations such as early detection with PSA and more aggressive treatment of both localized and metastasized disease.[16, 17] To what extent the reduction is explained by an effect of screening is debated. Modeling studies have indicated that up to 45-70% of the mortality reduction seen in the US could be attributed to screening and that changes in treatment could explain about a third of the reduction.[16, 17] In the USA, PC mortality has decreased by 45% since its peak in 1991.[18]

1.2.2 Natural course of prostate cancer

Prostate cancer is a heterogeneous disease where the natural course can range from latent, slow-growing disease to fast-growing and aggressive, leading to death within a couple of years. Knowing the natural course of untreated PC, is important in order to choose the optimal strategy for a man with newly diagnosed PC. However, this clinical presentation has changed since the

introduction of PSA-testing. Today most cancers are diagnosed at an early stage and a substantial proportion of screen-detected tumors are overdiagnosed.[19] Information regarding the natural course of PC can be obtained from several different sources.

Autopsy studies are one important source as they illustrate the true prevalence of PC and give an indication on the upper limit of the amount of PCs that could potentially be detected with screening. A recent review of autopsy studies of white men with no clinical diagnosis of PC during their lifetime reported that PC was detected at autopsy in 16% of men in their 50's and 40% of men in their 70's.[20] It is unknown how large a proportion of these latent autopsy cancers that are detected with screening and that proportion is probably dependent on factors such as PSA threshold and number of biopsy cores taken. Konety et al. reported that the detection rate of latent PC at autopsy decreased 3-fold since the introduction of PSA which could indicate that PSA-testing detects a proportion of these autopsy cancers.[21]

Information regarding the natural history of PC can also be obtained by observing a group of men who remain untreated. However, there are no "true" natural history studies for PC, even the observations studies that are generally referred to as the "natural history studies of PC" included men who received treatment, i.e. endocrine treatment for those with advanced disease. Despite these shortcomings, the studies by Chodak, Johansson, and Albertsen have contributed greatly to the understanding of the natural history of clinically diagnosed PC.[22-24] In 1994, Chodak et al presented a pooled analysis of 828 men from six non-randomized studies on deferred treatment for clinically localized PC. Men with well- and moderately differentiated tumors (cytological grade 1 and 2, corresponding to Gleason score ≤7) had a 10-year disease specific survival of 87% in comparison to 34% for those with poorly differentiated tumors.[22] The Johansson study consisted of 642 men with PC in Sweden who did not receive immediate treatment. Prostate cancer mortality was associated with grade of differentiation; 15-year PC mortality was 6% for highly differentiated tumors, 11% for moderately and 56% for poorly differentiated tumors.[23] Albertsen et al. used the Connecticut Tumor Registry to identify 767 men aged 55-74 who were diagnosed with localized PC between 1971 and 1984 and managed conservatively.[24] In the well-known Albertsen's tables he depicted that the risk or dying from PC was closely related to Gleason score (for description see paragraph 1.2.3) and age at diagnosis. Men with well-differentiated tumors (Gleason score <6) had a low-risk (<11%) of dying from PC within 15 years, whereas men with moderately differentiated PC (Gleason score 7) had a risk between 42-70%

and men with poorly differentiated tumors (Gleason score 7-10) had a high risk of dying from PC (70-87%) even if they were diagnosed as late as at age 74 years. [24]

The follow-up time is important when reading natural history studies. If followed long enough, localized PC that has remained relatively stable for 10-15 years can shift to become more aggressive and lead to metastasis and PC death. Aus et al. reported already in 1995, in a much debated article, that PC mortality for men with initially non-metastasized disease was eventually as high as 50% if the patient lived long enough.[25] In the Johansson cohort, which has an impressive follow up of more than 30 years, PC mortality increased three-fold after 15 years but remained relatively stable thereafter. The results from these studies could not be corroborated by Albertsen et al. who reported on 20 years of follow-up from the Connecticut Tumor Registry, in which there was no significant difference in PC mortality rate before and after 15 years of follow-up.[26]

The above mentioned studies were all conducted in the pre-PSA era and the study populations consisted mainly of men with palpable tumors (clinical stage T2). Therefore it has been questioned how the results can be applied to today's men with screen-detected, non-palpable tumors. Screen-detected tumors differ substantially from clinical tumors with respect to the PSA-level at diagnosis, grade and stage distribution.[7] Furthermore, an additional 3-12 years need to be added to survival estimates for screen-detected tumors due to *lead time*, i.e. the time that screening advances diagnosis as compared to its clinical presentation. Natural history studies of screen-detected cancers are scarce as the majority of men in the PSA era have been curatively treated. [27] One study by Lu-Yao et al. reported on the outcome of >14,000 men diagnosed between 1992 and 2002 at a median age of 78 years in the Surveillance, Epidemiology, and End Results Program (SEER)-database. A third of the population had T1c tumors and an additional third was T1a or T1b tumors. Ten-year PC-specific mortality was 8.3% for well-differentiated (Gleason score 2-4), 9.1% for moderately differentiated (Gleason score 5-7) and 25.6% for poorly differentiated tumors (Gleason score 8-10). The corresponding 10-year risks of dying from causes other than PC were 59.8%, 57.2% and 56.5% for each group respectively.[28] Explanations, other than lead time, for the improvement in survival for screen-detected cancer include overdiagnosis, grade migration and improvements in medical care.

The 2005 update of the Gleason grading system by the International Society of Urological Pathology (ISUP) has also had an effect on prognosis (see paragraph 1.2.3). A study by Gulati et al. used three different models to

project the risk of clinical detection, progression and PC death for screen-detected PCs in the absence of any treatment. These risks were closely related to age at diagnosis and Gleason score. For example, the risk that men diagnosed with a Gleason score 2-7 tumor before the age of 60 would have been diagnosed clinically in the absence of screening was 67-93% (depending on model) and the risk that they would have died from PC in the absence of treatment was 23-34%. For the same age group, but with a Gleason score 8-10 tumors, these risks were 90-96% and 63-83% respectively.[29]

Knowledge about the presumed natural history of screen-detected PC can also be gained from studies in which PSA has been measured in archived, frozen blood samples from men who later develop PC, by calculating the time from an elevated PSA-level to clinical diagnosis. These studies can also give estimates of lead time. For example, a study by Hugosson et al. compared the incidence and outcome of clinically detected PC in a group of men, from whom a venous blood sample had been drawn several years earlier, with the incidence and outcome of men with screen-detected PC in the first screening round of the Göteborg screening trial. With a PSA threshold of 3 ng/mL to perform sextant biopsies the detection rate of PC was 16.2% in the screening group (of those with PSA 3-10 ng/mL), which can be compared to the cumulative risk of developing clinical PC after 15 years which was 29% for men with a PSA in the range 3-10 ng/mL. The prognosis for those with moderately elevated PSA (3-10 ng/mL), who later developed clinically detected PC was relatively poor with a chance of survival at only 50%.[30] An often cited study based on stored serum samples is the study by Gann et al. which reported on 366 men in the U.S. Physicians Health Study who were diagnosed with palpable PC. For these men, PSA was elevated 4-6 years before diagnosis.[31] Several studies based on the Malmö Preventive Project, which was originally a cardiovascular study in the 1970s and 80s, have added further evidence to the strong association between PSA-levels in midlife and the future risk of clinical PC. For example, it has been shown that a PSA at age 44-50 can predict the risk of developing both palpable and advanced PC up to 20 to 30 years later. For men with the highest PSA-levels (≈1 ng/mL or higher) the absolute risk of palpable PC was 8-12% and the risk of advanced PC was 4-6% after a median follow-up of 23 years. Eighty-one percent of advanced PC were found in men with a PSA above the median (>0.63 ng/mL) at age 44-50 years.[32] The same group has also reported that PSA at age 40-55 years can predict the risk of PC metastasis and mortality. Although a PSA below median could not be used to rule out PC death within the coming 20-30 years, a PSA below median at the age of 45-49 or 51-55 years was associated with very low-risk of PC death within 15 years (0.09%

and 0.28% respectively).[33] PSA-level at age 60 years has also been shown to be predictive of the risk of dying from PC at age 85. Ninety-five percent of PC deaths occured in men with a PSA-level above the median (>1 ng/mL).[34] The studies based on stored serum samples show that there is a clear association between the PSA-level and the future risk of clinically significant PC.

Despite these studies the knowledge of the natural course of screen-detected PC is limited and additional research is needed in order to optimize screening as well as treatment of screen-detected cancers.

1.2.3 Grading, staging and risk groups

Prostate cancer is described by its grade and stage and these characteristic, together with the PSA-level at diagnosis are used for categorization of PC into different risk groups. Risk groups are predictive of prognosis and are valuable for guiding treatment decisions.

Grading: the Gleason score

In the 1960s, Donald F Gleason (1920-2008) created a grading system for PC based on the architectural pattern of the tumor.[35] It consisted of the sum (also called score) of the two most common patterns, called grades. Each of the grades could vary between 1 (well differentiated) and 5 (poorly differentiated), resulting in a sum between 2 and 10. It the most recent update by the International Society of Urological Pathology (ISUP) in 2005 it was decided that, for prostate biopsy material, pattern 1-2 should rarely, if ever, be used, and the Gleason score should instead be the sum of the most common grade plus the highest (worst) grade, yielding a summary Gleason score ranging between 6 through 10. For radical prostatectomy specimens, the Gleason score should still be reported as the two most common patterns but with a comment if small foci of higher grade were present.[36]

The changes in the reporting of the Gleason score has resulted in a relative upgrading of PC nowadays, as compared to the reporting done before 2005. This has led to an artificial change in prognosis, called the Will-Rogers phenomenon.[37] Many tumors that would have been graded as Gleason score 6 before the 2005 update would now be graded as Gleason score 7. This implies that the prognosis for men with tumors assigned today's Gleason score appears relatively better. This migration has been accompanied by an increased concordance between biopsy Gleason score and surgical Gleason score from 58% to 72%.[38]

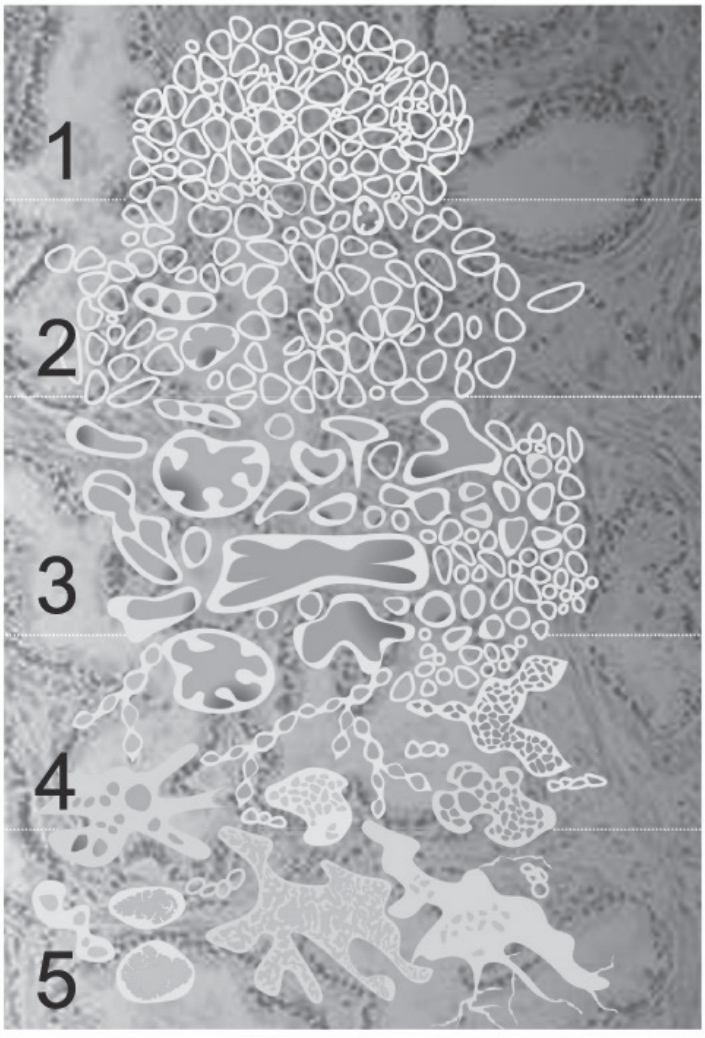

Figure 4. Schematic diagram of the Gleason grading system. (Reprinted with permission from AstraZeneca).[39]

Pathological grading is not an exact science; there is substantial intra- and inter-observer variability, even for pathologists specialized in urology.[40-42] The rumor has it that even Dr Gleason himself admitted that he duplicated his previous score about half the time. The European Randomized Study of

Screening for Prostate Cancer (ERSPC) is a large randomized trial of PSA screening conducted in eight European countries. This study has an international pathology committee. This committee performed a quality assurance on their own work and reported on the frequency of false-negative and false-positives biopsies by comparing the primary pathology reading from each ERSPC-center to that of 2 reference pathologists. For sextant biopsies false negatives occurred in 4-10% and false-positives 0.36% (cancer versus no cancer).[43]

The Gleason score has shown to be one of the strongest prognostic factors for the clinical behavior of PC as well as for treatment response. Gleason score in included in all nomograms (a multivariate prediction tool, see chapter 6.1 for further description) and risk prediction tools for PC.[28, 44]

Staging

The extent of the disease is commonly classified according to the Tumor Node Metastases (TNM) classification where T denotes the tumor size, N the extent of any lymph node involvement and M the presence or absence of distant metastases (Table 1).

Clinical staging can be performed with PSA, digital rectal examination (DRE) and transrectal ultrasound (TRUS).[11] A bone scan is performed if there is a risk of bone metastases. This examination is sometimes complemented with radiology exams, such as computed tomography (CT), magnetic resonance imaging (MRI) or positron emission tomography (PET-CT) in selected cases. Both PSA and DRE are poor predictors of the final stage at radical prostatectomy, but the combination of PSA, clinical T-stage and Gleason score at biopsy perform better than either variable alone. [45] MRI for T-staging has a high specificity (61-100%) for extracapsular extension of cancer but is limited by low sensitivity (22-82%) as MRI cannot detect microscopic extracapsular cancer growth. MRI can therefore be an alternative for T-staging only for selected intermediate and high risk cases.[11]

Examinations with respect for N and M stage should only be performed if the outcome of the examination will affect the treatment decision, for example in high risk patients when discussing curative treatment. The risk of lymph node involvement can be assessed with nomograms or Partin tables.[46, 47] MRI and CT are also alternatives for N-staging but are limited by low sensitivity (<40% for a 10 mm threshold) and cannot detect microscopic lymph node invasion. They are therefore not routinely recommended but can be an alternative in high risk patients.[48, 49]

T-primary tumor	
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
T1	Clinically inapparent tumor not palpable or visible by imaging
T1a	Tumor incidental histological finding in ≤5% of tissue resected
T1b	Tumor incidental histological finding in >5% of tissue resected
T1c	Tumor identified by needle biopsy (e.g. because of an elevated PSA
T2	Tumor confined within the prostate
T2a	Tumor involves one half of one lobe or less
T2b	Tumor involves more than half of one lobe, but not both lobes
T2c	Tumor involves both lobes
T3	Tumor extends through the prostatic capsule
T3a	Extracapsular extension
T3b	Tumor invades seminal vesicle(s)
T4	Tumor is fixed or invades adjacent structures other than seminal vesicles
N-Regional lymph nodes	
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Regional lymph node metastasis
M-Distant metastasis	
MX	Distant metastasis cannot be assessed
M0	No distant metastasis
M1	Distant metastases (M1a-non regional lymph nodes, M1b-bone, M1c-other site)

Table 1. The 2009 Tumor Node Metastases (TNM) classification for prostate cancer. (Adapted from [50])

The gold standard for N-staging is pelvic lymph node dissection, that is, surgical removal of the lymph nodes in the pelvis.[49] The exact extent of the lymph node dissection and for which patients it should be performed is debated. The 2014 European Association of Urology (EAU) Guidelines on Prostate Cancer recommends that limited lymph node dissection should not be performed as it misses >50% of nodes involved and that extended lymph node dissection is not necessary for low-risk patients but is indicated for some intermediate risk patients and for high-risk patients.[49]

M-stage investigations are usually restricted to a bone scan as the skeleton is the main site for distant metastases from PC.[51] The diagnostic performance of a bone scan is highly dependent on the PSA-level, Gleason score and clinical stage and bone scans therefore usually reserved for symptomatic patients, or alternatively, asymptomatic patients with a PSA > 10-20 ng/mL and/or Gleason score >7.[48, 52] Suspicious “hotspots” on a bone scan can be further evaluated with an MRI or CT.

Risk groups

Risk groups are used to provide information on prognosis and the risk of recurrence after treatment. There are several different definitions (Table 2). A recent publication from the NPCR, which reported on long-term mortality of men with non-curatively treated PC in Sweden, illustrated the importance of a risk group classification. For men diagnosed with PC from 1991 through 2009 PC mortality varied 10-fold according to risk group and age. For men younger than 65 years at diagnosis 15-year cumulative PC mortality ranged from 5.5% for those with low-risk PC to 81% for those with distant metastases at diagnosis. For the entire study population, 15-year cumulative PC mortality was as follows; low-risk 8.9%, intermediate risk 19.6%, high risk 35.5%, regionally metastatic 49.1% and distant metastases 69.5%.The degree of comorbidity according to the Charlson’s Comorbidity Index[53] was strongly associated with the risk of dying from competing causes especially for men <65 years. Competing causes of death were common across all age and risk groups. The 15-year cumulative risk of dying from competing causes ranged from 49.5% for those with low-risk PC to 27.7% for men with distant metastases at diagnosis. [54] These figures clearly show the variable natural history of PC and the importance of competing causes of death; many men with PC, even those with advanced disease, die from causes other than PC.

	Very low-risk	Low-risk	Intermediate risk	High risk	(Locally) advanced
D'Amico[55]		PSA ≤10 ng/mL and GS <7 and cT1-2a	PSA 10-20 ng/mL or GS=7 or cT2b	PSA >20 ng/mL or GS >7 or cT2c	
CAPRA score[56]		<3	3-5	6-10	
EAU[11]		PSA <10 ng/mL, GS<7, cT1c	PSA 10-20 ng/mL, GS 7 or cT2b-c	PSA<20 ng/mL, GS 8-10 or ≥cT3a	
NCCN[57]	PSA <10 ng/mL, GS<7, cT1c, PSAD <0.15, <3 cores with cancer	PSA < 10 ng/mL, GS < 7, cT1-2a	PSA 10-20 ng/mL, or GS 7, or cT2b-c	PSA > 20 ng/mL, or GS > 7, or cT3a	cT3b-4
NPCR[58]	PSA<10 ng/mL, GS ≤ 6, PSAD<0.15ng/mL, T1c, ≤4 cores with cancer, ≤8mm cancer in total	PSA<10 ng/mL, GS ≤ 6, T1-2	PSA 10-20 ng/mL, GS =7 and/or T1-2	PSA 20-50 ng/mL, GS 8-10 and/or T1-2	PSA >50 ng/mL and/or T3
ERSPC[59]		GS ≤6 and cT1-2	GS 7 and T1-2 or GS ≤7 and T3	GS 8-10 and T1-3 or T4 and any GS	PSA >100 ng/mL and/or M1
Göteborg screening study[2, 60]	PSA density <0.15 ng/ml, GS ≤6, T1c, <3 cores with cancer and ≤50% cancer in any core	PSA <10 ng/mL, GS ≤6, T1	PSA <20 ng/mL, GS ≤7 and/or T1-2	PSA <100ng/mL, GS ≥8 and/or T1-4	PSA ≥100 ng/mL and/or N1 and/or M1

*Table 2. Different risk group criteria *CAPRA, Cancer of the Prostate Risk Assessment; NCCN, National Comprehensive Cancer Network. GS=Gleason score.*

Risk groups are also predictive of the risk of biochemical recurrence, that is an increase in PSA-level after radical prostatectomy and/or radiation therapy, and the prognosis after treatment with curative intent, as was demonstrated in the studies by D'Amico et al.[55, 61] The D'Amico criteria are commonly used. They have been externally validated and shown to predict disease recurrence and survival following radical prostatectomy also in the more contemporary era.[62, 63] One limitation with risk group criteria, which becomes obvious when reviewing table 2, is that they vary between studies and organizations which make direct comparisons of estimates of outcomes difficult.

1.2.4 Treatment alternatives

Treatment for PC can have a curative intent or be symptomatic and palliative. Treatment strategies for PC include active surveillance, radical prostatectomy, various forms of radiotherapy, watchful waiting, hormonal treatment and a range of palliative chemotherapeutic agents for castration resistant PC.

Radical prostatectomy can be performed with an open retropubic- or, less commonly, perineal technique or as a laparoscopic procedure with or without robot-assistance. During the procedure the entire prostate gland (between the urethra and the bladder), some surrounding tissues and often also the seminal vesicles, are removed. A urinary catheter is placed and the urethra is reattached to the bladder neck. Whether or not to spare the neurovascular bundles (uni- or bilateral nerve-sparing procedure) depends on the size and localization of the tumor, the patient's age, preoperative potency status and patient preference. As previously mentioned, a limited lymph node dissection is no longer recommended and the indication for extended lymph node dissection is debated.[11, 64] However, it may provide important information about the prognosis but is associated with increased morbidity with complications including lymphoedema, lymphocele, deep vein thrombosis and pulmonary embolism.[65, 66]

Radical prostatectomy is one of few surgical procedures that is considered "evidence-based", i.e. there is level 1 evidence from a randomized controlled trial (RCT), the Scandinavian Prostate Cancer Group-4 trial, that this surgery provides a benefit in terms of both overall and disease-specific mortality compared to watchful waiting (observation). In this landmark study, the relative risk of PC death was 0.62 (95% CI, 0.44 to 0.87; P=0.010) with radical prostatectomy compared to watchful waiting, after 15 years of follow-up. The number needed to treat to avoid one PC death was 15 overall and 7

for men younger than 65 years.[67] The study population consisted to a large extent of clinically diagnosed cancers – only 12% were clinical stage T1c (non-palpable tumors). The results of SPCG-4 could not be corroborated in the PIVOT trial, where the study population consisted predominately of men with PSA-detected clinically localized PC (50% T1c tumors). After a median follow-up of 10 years, there was no significant difference in overall- or PC mortality between radical prostatectomy compared to observation (HR, 0.88; 95% CI, 0.71-1.08; p=0.22 and HR, 0.63; 95% CI, 0.36-1.09; p=0.09 respectively). However, a reduction in overall mortality was seen for men with PSA >10 ng/ml and possibly also for men with intermediate and high-risk disease. [68] At the time of writing this thesis there has been no high-quality randomized or prospective observational study that has been able to show that robot-assisted radical prostatectomy is superior to retropubic radical prostatectomy with regards to oncological and functional outcome. However, robot-assisted radical prostatectomy is associated with less blood-loss and lower transfusion rates as well as shorter hospital stay, but comes at a higher monetary cost.[69-74] Post-operative incontinence and erectile dysfunction are the two most common side-effects regardless of operative technique. The experience and skills of the individual surgeon have been shown to be very important for functional and oncologic outcomes.[75-77]

Radiotherapy can be given as external beam radiotherapy (EBRT), as low- (LDR) or high-dose rate (HDR) brachytherapy or as a combination or EBRT and HDR brachytherapy. Evolving techniques such as three-dimensional conformal radiotherapy and intensity-modulated external-beam radiotherapy have increased the precision of the radiation and therefore enabled dose-escalations up to 78-80 Gray (Gy). The standard for ERBT in Sweden today, is to give 78 Gy/39 fractions.[64] These higher doses have in several studies shown to provide superior long-term cancer control.[78, 79] LDR brachytherapy is only performed at a few clinics in Sweden. Radioactive seeds of iodine or palladium are permanently implanted in the prostate. During a period of months these seeds deliver a high dose of radiation to the gland with limited damage to the surrounding tissues. HDR brachytherapy is given in combination with EBRT. After a series of EBRT, the HDR brachytherapy is given by placing hollow needles filled with radioactive material (iridium). Results from observational studies suggest that the effect regarding cancer control for the different strategies of radiotherapy is comparable.[80-82] Radiation to pelvic lymph nodes is generally not recommended as randomized trials have failed to show a beneficial effect of this procedure.[83] Several studies, including the Swedish SPCG-7 trial have shown that radiotherapy combined with hormonal treatment is superior to radiotherapy alone for intermediate and high-risk PC.[84-87]. Therefore,

androgen deprivation therapy is recommended before and after radiotherapy for high-risk PC.[64] Neoadjuvant hormonal treatment before radical prostatectomy is not recommended as it deranges the histological picture and makes it difficult to decide on the need for adjuvant treatment. Radiotherapy can be given adjuvant to radical prostatectomy for those with a high risk of local recurrence or as "salvage radiation" after biochemical recurrence (2 consecutive PSA values of >0.2 ng/mL post-operatively).

Active surveillance is a relatively new treatment strategy, introduced in the last decade. It aims at reducing overtreatment of screen-detected PCs by postponing, or in some instances completely avoiding, curative treatment. In contrast to watchful waiting AS has a curative intent. Active surveillance implies that selected men with a low-risk PC are closely monitored with regular clinical examinations, PSA-tests and repeat biopsies. If there are signs of disease progression, treatment (radical prostatectomy or radiation therapy) is recommended with the intention to cure the patient. The challenge with active surveillance is to find the right candidates and the right signs of disease progression, so that the window of cure is not missed. Up to this day, there has been no RCT comparing active surveillance with other treatment strategies. The results from a large trial in the UK, the Prostate testing for cancer and Treatment (ProtecT)-trial, which has randomized men age 50-69 years with localized PC to radical prostatectomy, radiation therapy or active surveillance, are awaited in 2016. There are several large active surveillance-cohorts around the world which up to this date have reported medium-term outcomes (see subchapter 5.1 table 16 and 17). There are no evidence-based criteria for inclusion to active surveillance, how these patients should be followed or what should trigger intervention. Reported cohorts all have slightly different inclusion criteria but all are variations of very low and low-risk PC and generally include: clinically confined PC (T1-2), Gleason score <7, PSA <10-15 ng/mL and some volume criteria in biopsies. Some series also include selected men with intermediate risk cancer (see subchapter 5.1 table 15 and 16).

Watchful waiting is a strategy where symptomatic treatment (endocrine treatment, trans-urethral resection of the prostate (TUR-P), and/or radiation therapy for bone metastasis) is initiated only if the cancer progresses and causes symptoms. It was a common strategy in the pre-PSA era and before radical prostatectomy became a wide-spread procedure. It has palliative intent and is mainly a treatment alternative for men with limited life-expectancy due to age and/or a high degree of co-morbidity.

Endocrine treatment is generally reserved for advanced PC. The rationale behind endocrine treatment is that PC cells are dependent on testosterone for growth and perpetuation. The testes, controlled by the hormones LH and FSH, which in turn are controlled by the hormone LNRH, are the main sources of androgens but the adrenal glands also produce smaller amounts. Androgen deprivation can be achieved by suppressing the production by surgically removing the testes or by inhibiting LNRH secretion with agonists (negative feedback) or antagonists. The effect of androgens can also be blocked at the receptor level by anti-androgens.

Choice of treatment is determined by tumor risk group, the patient's general health, life-expectancy and patient preference. Various nomograms can be used to predict different PC outcomes.[88] The first assessment to make is whether the cancer is localized and potentially curable. Prostate cancer is believed to be beyond cure if there are distant metastases, lymph node involvement, if the PSA-level is above 100 ng/ml and if the tumor is clinical stage T4.[64] For these patients surgery or radiotherapy are not treatment options. In addition, treatment with curative intent is seldom recommended if the patient has a remaining life expectancy of less than 10 years, regardless of risk group. For men who are not candidates for curative treatment, the decision is when, or if, to start hormonal therapy. Hormonal treatment is generally recommended when symptoms arise. As there are no RCTs that have compared different curative treatments for PC with one another, the treatment decision is largely based on patient preference. Some cohort studies suggest that surgery is superior to radiotherapy in terms of oncological whereas other show similar results for the two treatment options.[89-92] Patients with very low-risk and low-risk tumors can be candidates for active surveillance. Surgery is the preferred alternative for those with higher risk tumors and larger prostates and radiotherapy is generally the first option for those with locally advanced tumors, clinical stage T3. The side-effects of various treatments and their effect on quality-of-life (QoL) are discussed in depth in paragraph 1.6.4.

1.3 Prostate-specific antigen

PSA is the most commonly used tumor marker in oncology. The early research on what would later be called PSA was performed by Flocks, Ablin and Hara in the 1960's and 70's. The initial research work on PSA in semen was carried out to assess PSA's properties as a forensic marker for rape victims. In 1979 Wang was the first to purify PSA from the prostate[93] and in 1987 Stamey reported that PSA was a more sensitive marker for PC than

prostatic acid phosphatase which had previously been used.[94] Work by several researchers in the early 1990's including Catalona, Labrie and Brawer showed that PSA could be used to assess risk of PC.[95-98] The U.S. Food and Drug Administration (FDA) approved the use of PSA for monitoring treatment response and disease recurrence and in 1994 it was approved as a screening tool for early detection of PC in asymptomatic men.[99]

PSA is a serine protease produced by the luminal cells of the prostatic epithelium. It is a member of the kallikrein gene family and is secreted in high concentration into the seminal fluid which is mixed with the semen during ejaculation. PSA lyses the protein semenogelin and this process liquefies the ejaculate, hence improving the motility of the sperm. PSA expression is strongly influenced by androgens and PSA starts becoming measurable at puberty. PSA is normally found in low concentrations in serum where it circulates in both free and bound form (bound to alfa-1-antichymotrypsin, alfa-2-macroglobulin and alfa-1-antitrypsin).[4] It is believed that disruption of the normal prostatic architecture with breakage of the basal cell layer and the basal membrane, caused by for example inflammation, BPH, PC or trauma (i.e. prostate surgery, biopsies, cystoscopy and urine retention) allows PSA to leak out to the circulation where it can be measured as a PSA elevation.[4] The characteristics of PSA as a screening test for PC are discussed in depth in chapter 1.5.1.

1.4 Screening

1.4.1 Definitions and strategies

Screening, as defined by the World Health Organization, is "the use of simple tests across a healthy population in order to identify individuals who have a disease but do not yet have symptoms".[100] It is a form of secondary prevention, which aims at finding individuals who are likely, or unlikely, to have the disease of interest so that the individual can be the subject to further work-up. The concept of screening is based on the assumption that early diagnosis and treatment will improve prognosis as compared to later diagnosis and treatment at a symptomatic stage. The main goal of cancer screening is to reduce disease-specific mortality. Screening can be performed with three different strategies: mass screening (screening the entire population), selective screening (e. g. screening only high-risk individuals), or opportunistic screening (non-organized screening, e. g. an individual having a certain test as part of a laboratory work-up performed in clinical practice or a test as part of a physical examination).[101]

According to the International Agency of Research of Cancer (IARC) an organized screening program is defined by [101]:

- a written policy stating the target population, method of screening and screening interval
- a defined target population
- a management team responsible for implementation
- a health care team responsible for decisions and care
- a quality assurance structure
- a method for detecting cancer occurrence in the target population

1.4.2 Characteristics of a suitable disease

To be suitable for screening a disease must have a preclinical (asymptomatic) phase when it can be detected by a screening test, and in addition, be such that is progresses over time. The preclinical phase starts with the biological onset of the disease. At some point during this process the disease becomes theoretically detectable by the screening test and the preclinical detectable phase starts. This point in time depends both on the characteristics of the disease and the characteristics of the test. The proportion of a population being in the preclinical detectable phase depends on the incidence of the disease, the average duration of the preclinical phase for that disease and any prior screening activity in the population.[102] If repeated screening takes place, the prevalence of the preclinical phase is determined mainly by the incidence of the disease. Generally, a disease is thought to be suitable for mass screening if the preclinical detectable phase is long enough and the prevalence of the preclinical detectable phase is high enough. In addition, the disease should have sufficient public health importance. As previously mentioned, early treatment must result in better prognosis and reduce disease-specific mortality, in comparison to diagnosis when symptoms arise, for screening to be effective.[101]

1.4.3 Characteristics of a suitable test

A suitable screening test must be cheap, ubiquitous in the population and acceptable for the individual, i.e. reasonably free from discomfort and risks. It must also be reliable and valid. Reliability of a test is its "capacity to give the same results – positive of negative, whether correct or incorrect – on repeated application in a person with a given level of disease".[102] Factors that affect reliability are the biological variability of the disease being screened for, the variability of the equipment and the intra- and inter-observer variability.[103]

Validity is the ability to correctly identify those who have and those who do not have the disease. The validity can be expressed by the test's sensitivity and specificity (Table 3).[101, 103] Sensitivity is the probability to correctly classify people who have the disease. It is often difficult to estimate sensitivity of screening tests, since as a definitive, confirmatory, diagnostic test to find all those who have the disease (cells a+c in table3) is usually not performed on the screening population. One strategy of dealing with this limitation is to follow all individuals and to observe the number of cases that develop among the screening negative and to regard them as false-negatives (interval cancers, i.e. cancers that are detected between screening intervals). However, it is often difficult to know whether these are "true" false-negatives or new cases of fast growing cancers. Specificity of a test is the probability to correctly classify individuals without the disease. An optimal screening test would have 100% sensitivity and 100% specificity which is never the case. The results of the screening test for individuals with and without disease often overlap if the screening test is continuous, for example PSA and the cut off value for positive versus negative test result has the be somewhat arbitrary chosen. In reality, there is often a trade-off between sensitivity and specificity.[101]

Moreover, a screening program also has to be feasible. Two estimates of feasibility are the positive predictive value (PPV) and the negative predictive value (NPV).[101, 103] PPV is the proportion of individuals who tests positive who have the disease and NPV is the proportion of individuals who test negative who are disease-free. A high PPV is important for a screening test to be considered useful as it indicates that the test detects many cases among those with a positive test and that diagnostic resources for further work up are not spent unnecessarily on disease-free individuals. PPV and NPV for a given screening test are influenced by sensitivity, specificity and disease prevalence and are therefore not fixed characteristics of a test. Specificity has a greater effect on PPV than sensitivity, but the opposite is true for NPV. Specificity determines the number of false positives and as most people in a population are disease-free, a small decrease in specificity might have a large effect on the absolute number of false positives, which can results in large drop in PPV. PPV also increases with sensitivity and prevalence. A screening program can therefore improve its PPV by focusing on high-risk individuals (higher prevalence).[101, 103]

	Truth/gold standard		
Test result	Disease	No disease	
Positive	**a** true positive	**b** false positive	**a/a+b**=PPV
Negative	**c** false negative	**d** true negative	**d/c+d**=NPV
	a/a+c=sensitivity	**d/b+d**=specificity	

Table 3. Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV)

In 1968 Wilson and Jungner (then Chief of the Clinical Chemistry Department at Sahlgrenska sjukhuset) wrote a report entitled "*Principles and practice of screening for disease"* for the WHO in which they summarize ten criteria that need to be met in order for a disease to be suitable for screening. These criteria have become classic and are often cited.[104]

Wilson and Jungner classic screening criteria

- The condition sought should be an important health problem
- There should be an accepted treatment for patients with recognized disease
- Facilities for diagnosis and treatment should be available
- There should be a recognizable latent or early symptomatic stage
- There should be a suitable test or examination
- The test should be acceptable to the population
- The natural history of the condition, including development from latent to declared disease, should be adequately understood
- There should be an agreed policy on whom to treat as patients
- The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole
- Case-finding should be a continuing process and not a "once and for all" project

As screening generally is directed towards healthy people, the medical and ethical demands must be higher than for diagnosis and treatment for individuals with disease signs and symptoms. As the field of medicine has evolved with an increasing focus on evidence-based medicine, patient autonomy, cost-effectiveness and quality control several adaptions to the classical criteria have been suggested. Andermann et al., also on behalf of the WHO, published ten new criteria, as a summary of the requirements that have emerged during the last 40 years.[105]

Andermann's revised screening criteria

- The screening program should respond to a recognized need
- The objectives of screening should be defined at the outset
- There should be a defined target population
- There should be scientific evidence of screening program effectiveness
- The program should integrate education, testing, clinical services and program management
- There should be quality assurance, with mechanisms to minimize potential risks of screening
- The program should ensure informed choice, confidentiality and respect for autonomy
- The program should promote equity and access to screening for the entire target population
- Program evaluation should be planned from the outset
- The overall benefits of screening should outweigh the harm

1.4.4 Evaluation of screening

The main goals of a screening program are to reduce morbidity and mortality of the disease being screened for. Other important measures are process measures, such as the number of persons being screening and the number of cases detected. Intermediate measures that can be evaluated before mortality data is available are stage migration and case fatality. Stage migration implies that screen-detected cases are diagnosed at an earlier stage than clinically diagnosed cases and is an indication of screening effectiveness. Case fatality is the number of deaths among cases.[101] The effectiveness of a screening program can be evaluated with different study designs such as experimental (e.g. a RCT), cohort, case-control and ecological studies.[102, 103]

There are several forms of biases that must be considered when evaluating screening. If survival is evaluated, a phenomenon called lead time bias may

occur. Lead time is the time by which screening advances diagnosis, i.e. the time from screen-detection to when the disease would have appeared clinically (Figure 5).[102] If survival is measured from time of diagnosis, survival of men with screen-detected cancer will appear longer, even if they do not actually live longer simply because they are diagnosed earlier. Another form of bias is length time bias, which is a form of selection bias. This can occur because screening tends to diagnose slow-growing, less aggressive tumors with a long preclinical detectable phase rather than the fast-growing and potentially lethal tumors (Figure 6). Volunteer bias is a third form of bias which occurs if people who volunteer for screening differ from those who do not.[103] Lead time and length time biases can be controlled for if screening is evaluated in RCTs of sufficient size where patients are randomized to screening versus no screening, where mortality is studied rather than survival, where all outcomes are counted regardless of method of detection (screen-detected and interval cancers) and where outcomes are compared for the entire study population and not just the participants.[102, 103]

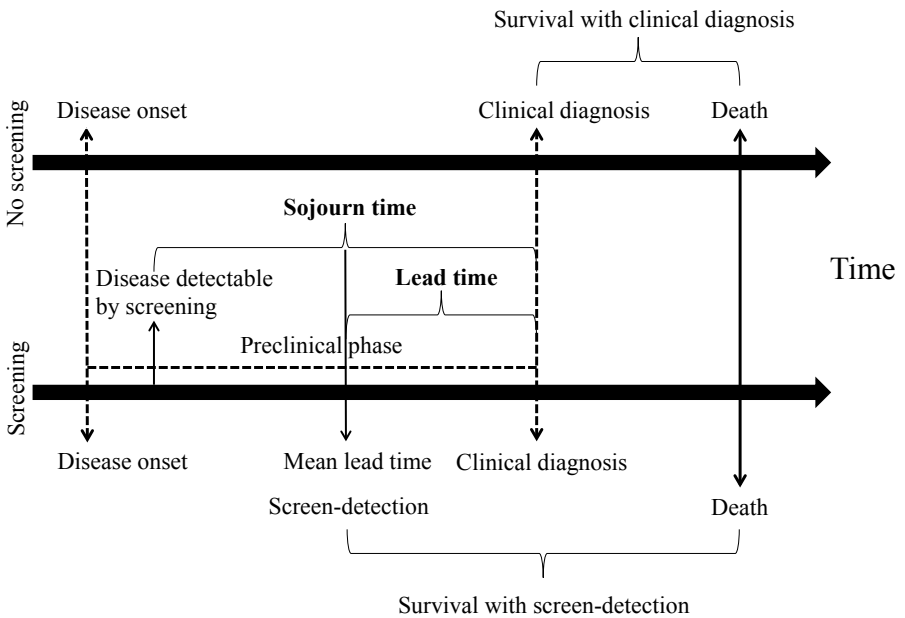

Figure 5. Lead time

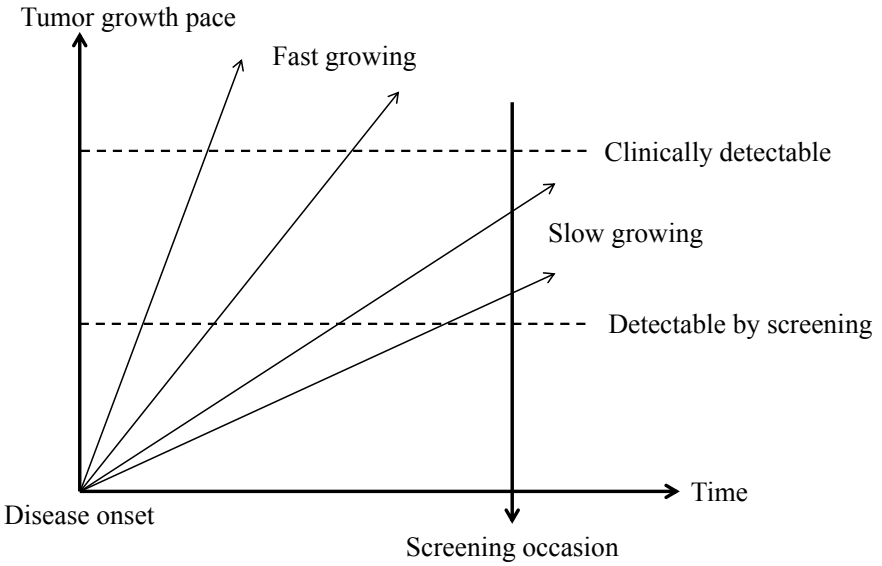

Figure 6. Length time (Adapted from [106])

When a cancer screening program is introduced in a population, the observed cancer incidence is expected to increase as screening advances diagnosis (lead time) but also because in the prevalence screen (the first screening round) tumors are detected from a pool of small, preclinical and sometimes latent tumors. When lead time is over, cancer incidence should drop to a level below the predicted incidence without screening if no overdiagnosis is present. This is because screening has advanced the time of cancer diagnosis. However, in a situation with overdiagnosis, the observed incidence will not fall below the expected but remain at a new and higher level. Overdiagnosis can, as previously mentioned, be defined as detection of a tumor with screening, which in the absence of screening, would never have been diagnosed. If a screening program is directed towards finding pre-cursor lesions, e.g. cervical cancer screening, the cancer incidence should decrease after the introduction of this screening program.

1.5 Screening and diagnosis of prostate cancer

Whether or not to screen for PC is one of the most controversial issues in urology. There is evidence from high-quality studies that organized PSA screening reduces PC mortality [2, 59] but opponents argue that the harms of screening outweigh the benefits.[107, 108] The recommendations from

various organizations therefore differ. The Swedish National Board on Health and Welfare do not recommend population-based screening for PC but states that PSA-testing should be an individual decision and that well-informed men who wish to have the PSA-test should be given this possibility after receiving written information.[109] The European Association of Urology (EAU) recommends an individualized, risk-adapted strategy where well-informed men with a life-expectancy of at least 10-15 years can be offered a PSA-test. They recommend the baseline PSA-level be used to determine the screening interval.[11] There are also several different American organizations that have issued screening guidelines. For example, the American Association of Urology (AUA) recommends shared-decision making about screening for men 55-69 years and recommends against screening for men <40 years and against routine screening for men 40-54 and >70 years. They also state that a screening interval of ≥2 years may be preferred but that a baseline PSA can be used to determine the screening interval.[110] The National Comprehensive Cancer Network (NCCN) recommends that healthy, well-informed men 50-70 years of age be offered screening and that screening above the age of 70 years should be individualized.[111] The US Preventive Services Task Force (USPSTF) updated their recommendation on PSA screening in 2012 and came to the controversial conclusion to give a Grade D recommendation *against* PSA screening for men of all ages.[107] Their recommendation has been heavily criticized.[112, 113] In summary, most organizations recommend shared-decision making for men with at least a 10-year remaining life-expectancy but recommendations regarding starting age and screening interval differ.

1.5.1 Screening tools

Digital rectal examination

DRE has long been used as a screening tool for PC, and it was the only screening method available before the advent of PSA. The examination has the advantages of being a simple, quick and cheap method without serious side-effects. However, the examination is subjective and the test performance is dependent on the skill of the examiner. Moreover, many tumors, even life-threatening, will never be palpable. Tumors in the peripheral zone (the majority of the tumors) are said to be palpable when the volume exceeds 0.2 mL.[11] Screening with DRE alone has never been investigated in a randomized trial but several studies have reported that the combined use of PSA and DRE increases detection rate.[98, 114, 115] For example, in the Dutch center of the ERSPC the PPV of suspicious DRE in conjunction with an PSA above 3 ng/mL was 49% compared to a PPV of 22% for those with a normal DRE but an elevated SPA.[115] An abnormal DRE has also been

shown to be associated with a more aggressive cancer (Gleason score ≥7) and should therefore prompt prostate biopsy.[115-117]

Prostate-specific antigen

PSA is, as previously mentioned, the most commonly used tumor marker in oncology but as a screening test it is far from perfect. Nevertheless, it is currently the best screening test available. Although PSA can be found in small amounts in certain cancers and tissues other than the prostate gland, there is no other significant source of serum PSA. Thus, PSA can be considered prostate-specific, from a clinical perspective. PSA elevations often indicates a prostatic disease but are not cancer-specific.[4] There is an overlap in PSA-levels between healthy men, men with BPH and those with PC. In BPH the amount of PSA is proportional to the prostate size. Prostate cancer cells, on the other hand, produce less PSA than normal prostate cells but the PSA-level is proportional to the number of cancer cells. Men with metastasized PC can therefore have PSA-levels reaching several thousands.

PSA is measured in a venous blood sample, and like any other laboratory test, reliability is affected by measurement errors and variations between different laboratory assays. The variation between different laboratories in Sweden is approximately 6% and measurement errors within each method range between 2-6%. There is also an intra-individual variation, and PSA-level can vary by 15% within a few weeks also for a person without prostatic disease.[64] It is therefore recommended that a man have a second PSA test within a couple of weeks after an elevated test.[118]

Estimating the sensitivity, specificity, PPV and NPV for PSA is difficult since not all men will undergo a biopsy and prostate biopsies are not the perfect "gold standard" but have a certain level of false negatives (depending on the number and location of biopsy cores). Also, the test performances of PSA depend on laboratory assays and methods, PSA threshold, number of biopsy cores, PC prevalence and whether or not overdiagnosed cancers are included. Cancer detection rate (number of cancer/number of men screened), however, is relatively easy to calculate, but is also dependent on both test and population characteristics. For example, in the Göteborg screening study, the detection rate in the first screening round for a PSA threshold 3ng/mL was 2.5%.[119]

There are several different approaches to estimate the sensitivity of PSA. From the Finnish arm of the ERSPC Auvinen et al. estimated sensitivity with the incidence method where sensitivity was calculated by comparing the incidence of cancers in screen negative men (interval cancers) to the

incidence in the control arm (expected incidence in the absence of screening). Men who were diagnosed with interval cancers were regarded as false negatives. With this method, test sensitivity for PSA >3 ng/mL was 89%.[120] Another approach was used in the Prostate Cancer Prevention Trial (PTCP).The PCPT was conducted to study the use of the 5-alfa-reductase inhibitor finasteride for the prevention of PC, and randomized men with PSA <3 ng/mL and normal DRE to either finasteride or placebo. The trial protocol included annual PSA and DRE, and a biopsy was recommended in case of PSA > 4.0 ng/mL and/or a DRE suspicious of cancer. In addition, at the end of the trial, all participants without PC were biopsied. In this low-risk population (all had PSA <3 ng/mL and normal DRE at entry), PSA sensitivity for a cut off of 4.1 ng/mL was 20.5% and specificity was 93.8%. The PCPT clearly showed that PSA could not be treated as a dichotomous variable; instead, there was a continuum of risk for PC at all PSA-levels.[121] Prostate cancer was found among 15% of those who had never had a PSA above 4 or an abnormal DRE during the study, and 15% of these tumors were Gleason score ≥7.[122] A third approach to estimate sensitivity was used by Gann et al. In an unscreened population, using stored serum samples, sensitivity for clinically detected PC within a 4-year period for a PSA cut-off of 4 ng/mL was 73%.[31]

Specificity has also been estimated in the Finnish section of the ERSPC for a PSA cut-off of 4 ng/mL and an ancillary test (DRE or free-to-total, F/T PSA) for PSA 3.0-3.9 ng/mL. Specificity was calculated as the proportion of screen-negative (true negative) men among those who were regarded as disease-free (true negative + false positive). False positives were those with an elevated PSA and/or abnormal DRE but no cancer in subsequent biopsies. With this method specificity for PSA > 4 ng/mL was 93% in the first screening round and essentially the same in the second round. Specificity decreased slightly with age.[123] One limitation with this method is that it assumes that men with a PSA below the threshold do not harbor cancer. As they were not biopsied, and even if they were, this cannot be known for sure. As with sensitivity, specificity can be calculated using stored serum samples. These studies, which were conducted in the pre-PSA era, have the advantage of being uninfluenced by overdiagnosis as they only included clinical cancers. Similar estimates of specificity have been reported from such studies (91%).[31]

The American Cancer Society performed a systematic review of the literature assessing PSA performance and reported pooled estimates for PSA thresholds of 4 ng/mL and 3 ng/mL (Table 4). In their calculation they also assumed that screen negatives were true negatives.[124]

	>4 ng/mL	>3 ng/mL
Cancer detection rate	3%	4%
Sensitivity	21%	32%
Specificity	91%	85%
PPV	30%	28%

Table 4. Pooled estimates for PSA performance (Adapted from [124])

These figures clearly show that there is balance between sensitivity and specificity and if the threshold for biopsy is lowered more cancers will be diagnosed but at the cost of more false positives (decreased specificity) leading to unnecessary biopsies. False positives are common, as indicated by the low PPV of PSA; most men with a PSA above 3 or 4 ng/mL do not have cancer. In the Göteborg screening study 24% of those with a PSA above 3ng/ml had PC at biopsy (PPV).[119] The NPV of PSA was estimated in the PCPT and for a PSA-level of ≤4 ng/mL, NPV was 85%.[122]

This suboptimal specificity of PSA, especially among men with moderately elevated PSA-levels (4-10 ng/mL), has led to a search for new ways of trying to improve upon the performance with i.e. PSA-density, F/T PSA and PSA kinetics. PSA density relates the PSA-level to the volume of the prostate measured by TRUS. As PC cells express more PSA per volume than BPH, a PSA density below 0.1-0.15 ng/mL/cm^3 is believed to be more likely to be indicative of BPH rather than PC.[125, 126] The ratio of F/T PSA has the advantage that it is not user-dependent as no TRUS is required. The proportion of free PSA is lowered in men with PC, and the lower the ratio the greater the risk of PC.[127, 128] The F/T PSA ratio is generally used together with PSA density and PSA kinetics in men with elevated PSA to determine indication for biopsy. PSA kinetics are measures of the change in PSA over time and can be calculated in a number of ways; as PSA velocity, PSA doubling time or percent change. Lately it has been questioned whether PSA-kinetics add any predictive information compared to PSA alone in predicting risk of low-risk PC on biopsy.[129, 130] A review article by Vickers et al. assessed the evidence from 87 published articles on PSA kinetics and

concluded that there was little evidence that PSA velocity or doubling time adds any predictive information in predicting risk of for untreated patients beyond that provided by the absolute PSA-level alone.[129]

Transrectal ultrasound and prostate biopsies

TRUS is a rapid and generally well-tolerated procedure. The classic image of PC on TRUS is a hypoechoic area in the peripheral zone of the prostate, however the sensitivity for such a lesion being PC is only 25-30%.[131] TRUS cannot be used as a primary screening tool because of a limited sensitivity and specificity for PC but has an important role in the visualizing the prostate and the seminal vesicles, in measuring prostate volume and in guiding the biopsy needle.

TRUS-guided prostate biopsy is the clinical standard for a definitive PC diagnosis. Biopsies with a transrectal approach are most common but they can also be performed transperineal. During the transrectal approach an ultrasound-guided periprostatic block of local anesthetics is given to reduce procedural pain and antibiotic prophylaxis, commonly with ciprofloxacin, is given to reduce the risk of infectious complications. During the last 30 years, there has been a strong trend of increasing the number of biopsy cores sampled; from four cores in the early 1980's, to laterally directed sextant biopsies which was the standard for many years, to around 12 cores today. However, because sextant biopsies miss up to 34% (false negatives) of all PC and since the cancer detection rate of 10-12 cores is approximately 30% higher than sextant biopsies,[132] the current EAU and AUA recommendation are that a minimum of 8-12 cores be sampled.[11, 133]

Several studies have reported on the limitations of TRUS-guided biopsy in accurately grading PC. This is crucial as Gleason score is important for proper treatment selection and prediction of PC prognosis. For instance, for men being considered for active surveillance, upgrading on an early repeat biopsy occurs in approximately 20-30%.[134-136] This is likely a reflection of under sampling at the first biopsy rather than true disease progression, considering the long natural history of PC. A proportion of men eligible for active surveillance also has a negative biopsy or down-grading (20-45%) on their repeat biopsy which further indicates that the biopsy results are not a perfect reflection of the tumor burden but only mirror small parts of the prostate.[134, 136] The concordance between biopsy results and the patoanatomical review of prostatectomy specimens is highly dependent on the number of biopsy cores taken. The concordance rate for extended biopsy schemes (10-12 cores) range from 52% to 80% compared to 41% to 63% for sextant biopsies.[133] Various extended biopsy schemes such as saturation

biopsy (32 cores) or template guided mapping biopsies, where the entire prostate gland is covered with biopsies at 5 mm interval, have therefore been proposed, for example when selecting candidates for active surveillance.[137-139] The exact number of cores and their location is an on-going controversy. It is a balance between an adequate cancer detection rate with a high pathology concordance versus low levels of overdetection and side-effects at a reasonable cost. In a comprehensive review by Eichler et al. it was concluded that biopsy schemes with 12 cores (a sextant biopsy plus laterally directed cores) detected 31% more cancer than sextant biopsy, but increasing the number of cores to 18-24 did not detect significantly more cancers than 12. There was no difference in adverse events between sextant and 12 core biopsies.[140]

1.5.2 The evidence for prostate cancer screening

The question of whether or not to screen for PC with PSA has for years been heavily debated and is still controversial. Many hoped that the controversy surrounding PSA screening would be settled when the long-awaited first mortality reports from the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (the prostate section of this trial will hereafter be referred to as PLCO) and ERSPC were published in 2009, in the same issue of the *New England Journal of Medicine*. On the contrary, the debate became even more polarized as the two studied reported contrasting results. In addition, in 2012 the USPSTF published their Grade D recommendation *against* PSA screening and the 2013 Cochrane metaanalysis came to the conclusion that PSA screening did not significantly reduce PC mortality.[107, 108] This conclusion was reached by pooling the results of five very different studies. Below follows a short summary of these five randomized and the Göteborg screening trial which was not included in the analysis.

The Quebec trial

The Quebec trial randomized more than 46 000 men aged 45-80 years to annual screening with PSA and DRE versus no screening. A PSA above 3 ng/mL and/or a positive DRE were regarded as positive screening tests and led to TRUS-guided biopsies. A 62% reduction in PC mortality was reported after a median of 8 years.[141, 142] However, the data analysis has been heavily criticized as it did not follow the intention-to-screen principle; only men who actually were screened (23.6% of those randomized to screening) were compared to those who were not screened regardless of allocation arm at randomization. It is therefore difficult to assess the value of these reported results.

The Norrköping trial

The Norrköping trial was initially designed as a feasibility study of PC screening. Of 9026 men aged 50-69 years in Norrköping, Sweden, every sixth man was selected to a screening group (n=1494) and the remaining population constituted a control group. Men were screened every third year; during the first two rounds only with DRE and thereafter PSA was added. A suspicious DRE and/or a PSA above 4.0 ng/mL led to a recommendation of a sextant biopsy. After a total follow-up of 15 years, more PCs were detected in the screening group than amongst controls and screen-detected tumors tended to be more localized, but there was no significant difference in PC mortality between the groups.[143]

The Stockholm trial

In the Stockholm trial, 2400 men 55-70 years were randomized to a single screen with PSA, DRE and TRUS and the remaining source population (n= 24,202 men) constituted controls. A PSA >10 ng/mL and/or abnormal findings on DRE or TRUS led to quadrant biopsy. After a median of 12.9 years there was no difference in PC or overall mortality between the screening and control group. However, non-attendees in the screening group had a higher overall mortality than the attendees. The increased risk of death among non-attendees was due to non-PC mortality.[144]

The PLCO trial

The PLCO trial is a large randomized controlled cancer screening trial conducted in 10 different centers in the USA.[145] The prostate section of the trial randomized more than 76 000 men in a 1:1 ratio to screening or control. The screening group in the PLCO was offered annual DRE and PSA tests for 4 years and only PSA for another 2 years. A PSA >4ng/mL and/or an abnormal DRE led to recommendations for further diagnostic evaluation which was not part of the study protocol but was at the discretion of the patient and his primary physician. After 10 years of follow-up more PC was diagnosed in the screening group than the control group (rate ratio 1.22; 95% CI 1.16-1.29).[146] There was no sign of stage-shift between the groups; the number of men with in advanced disease was similar. There was no statistically significant reduction in PC mortality between the groups at 10 years. Fifty men died due to PC in the screening group compared with 44 among controls resulting in a rate ratio of 1.13 (95% CI 0.75-1.70).[146] Results after 13 year of follow up have also been published. At 13 years, the relative risk of PC diagnosis between the groups had decreased to 1.12. Again, there was no statistically significant reduction in PC between the groups (rate ratio 1.09, 95% CI 0.87-1.36).[147]

The ERSPC trial

The ERSCP was initiated in the 1990's to evaluate the effect of PSA screening on PC mortality. Belgium, Finland, France, Italy, the Netherlands, Portugal, Spain, Sweden and Switzerland were initially participating but Portugal discontinued their participation in 2000 and France did not join until 2001, hence the French data were not included in the 2009 publication. A total of 182160 men 50-74 years old underwent randomization of which 162 387 were in predefined core age group of 55-69 years. Recruitment and randomization procedures differed between countries following national rules and legislation. Men in the screening arm (n= 82 816) were invited for screening with PSA every fourth year (Sweden every second year). The screening algorithm varied somewhat between centers and could include DRE, TRUS and F/T PSA ratio as ancillary test but most centers used ≥3.0 ng/mL as the PSA cut-off for biopsy. Results have been published for 9-, 11- and 13-years of follow-up. After 9 years of follow up, the cumulative PC incidence in the screening and control group was 8.2% and 4.8% respectively. Of those randomized to screening 82% attended at least once and the average rate of compliance with biopsy recommendations was 86%. The incidence of localized PC was higher in the screening group while the incidence of metastasized disease was higher in the control group. In the core age group there were 214 PC deaths in the control group and 326 PC deaths in the control group at a median follow-up of 9 years. The rate ratio for PC death was 0.80 (95% CI, 0.65 to 0.98; P = 0.04). In order to prevent one PC death 1410 men needed to be invited to screening (NNI) during 9 years and 48 men needed to be diagnosed (NND) with PC.[148] In 2012, updated results with 11-years of follow-up were published. The rate ratio for PC incidence was 1.63; 95% confidence interval (95%CI 1.57-1.69). A relative reduction of 21% in PC mortality was reported (0.79; 95% CI, 0.68 to 0.91; P=0.001) which corresponded to a NNI of 1055 and NND of 37.[149] The recently published 13-year follow-up confirmed previous results that screening with PSA substantially reduces PC mortality. The rate ratio for PC incidence between screening and control arm had decreased to 1.57 (1.51-1.62) after 13 years and the rate ratio for PC mortality remained stable at 0.79 (95% CI 0.69-0.91). This resulted in improved estimates of NNI and NND which after 13 years of follow-up were 781 (95% CI 490-1929) and 27 (95% CI 17-66) respectively.[59]

The PLCO and the ERSPC have both been criticized, but for separate reasons. The criticism of PLCO has been focused on the following points:

-the population was heavily pre-screened; approximately 44% in each arm had at least one PSA-test before study start.[146]

-there was a high degree of contamination (PSA-testing) in the control group; after 6 years, 52% of the control group had been screened with PSA.[150]

-in the screening group only 40% of those with a PSA above threshold were in fact biopsied.[151]

It has been questioned whether the study had enough power to find a difference in PC mortality at these levels of contamination and compliance.[152]

Those who have criticized the ERSPC have argued that it is a multicenter study with different populations and different screening algorithms and that the distribution of different treatments has not been the equal between the study arms.[153]

The Göteborg trial

The Göteborg randomized population-based prostate cancer screening trial, which constitutes the study base for the four papers in the present thesis, was initiated in 1994.[2] It was planned as an independent study and became part of the ERSPC in 1996, with men in the core age group 55-69 years at randomization, without any changes in the protocol. As of today, the Göteborg trial is the study that has reported on the largest reduction in PC mortality due to PSA screening. The study population and screening algorithm if the Göteborg screening trial is described in depth in chapter 3.The first mortality results of the Göteborg screening study were published in the *Lancet Oncology* in 2010. Of the 9952 men in the screening arm, 76% attended at least once and among those with a PSA above threshold, 93% had at least one biopsy. After 14 years of follow-up the cumulative incidence of PC was 12.7% in the screening arm and 8.2% in the control arm, corresponding to a hazard ratio of 1.64 (95% CI 1.50-1.80). There was a stage migration with more low-risk cancers detected in the screening arm and more metastasized cancers in the control arm. At 14 years, a total of 44 men had died due to PC in the screening arm compared to 78 men in the control group. The rate ratio of PC death between men in the screening arm and controls was 0.56 (95%CI 0.39-0.82). NNI was 293 and NND 12.

1.6 Harms of prostate cancer screening

The introduction of PSA as a screening tool for PC has led to a substantial increase in the number of men diagnosed with PC. In addition, a substantial proportion of these cancers would never have caused symptoms or death. If organized PSA screening were to be introduced, even more PCs would be

detected. The mortality benefit of an organized screening program has to be weighed against a number of negative effects. The potential harms of screening are present in all stages of the screening process and further downstream following curative treatment for PC. Overdiagnosis and its consequences are, by many, regarded the primary harms

1.6.1 Lead time and overdiagnosis in prostate cancer screening

Lead time and the rate of overdiagnosis are closely connected. Lead time, as previously described, is the time that screening advances diagnosis, from screen-detection to when the tumor would have been clinically diagnosed. Overdiagnosis can be viewed from different perspectives; one can adopt an epidemiological view and define overdiagnosis as detection of a cancer that in the absence of screening would never have been diagnosed.[19] With this definition, it is impossible to know if a patient is overdiagnosed at the time of treatment; the only way to find out would be to leave the cancer untreated and follow the patient for his remaining life-time, and observe what he would die from. For obvious reasons, this is rarely an alternative. Therefore, this definition is not particularly helpful when trying to decide on the right treatment strategy for a patient. Another, more clinical way of looking at overdiagnosis is trying to identify tumor features that determine prognosis and to define tumors as clinically significant or insignificant. Overdiagnosis is then all cancers defined as clinically insignificant.

Most would agree that a 55-year old man with a large Gleason 4+5 tumor have a clinically significant tumor but if he would end up dying in a car crash soon after diagnosis, from an epidemiological perspective he could be regarded as overdiagnosed.

With either definition, overdiagnosis is a major problem in PC screening. A large study based on the nationwide SCR and the US SEER Program found that PC accounted for 52% of all deaths among PC patients in Sweden compared to 30% in the United States during the essentially the same time period (1961-2008 in Sweden and 1973-2008 in the US). Among men with a PC diagnosis the 10-year cumulative risk of dying from PC was 43% in Sweden and 20% in the United States. [154] The advent of PSA has led to a stage shift, and today most men are diagnosed with T1c tumors with very-low or low-risk features with excellent long term prognosis.[28]

Lead time cannot be measured directly for a single case but the lead time distribution can be estimated. The maximum lead time is the sojourn time

(duration of the preclinical detectable phase). It is assumed that on average the lead time is half of the sojourn time if the cancer is detected in the first screening round and shorter if the cancer is detected in subsequent rounds. Lead time distribution in a screening study can be estimated by comparing the frequency of diagnosed cases in the screened and unscreened population where the lead time distribution is the area between the two curves. The lead time distribution depends on the sensitivity of the screening test, screening interval, incidence of the disease and the duration of the preclinical detectable phase.[102]

Lead time can also be viewed as a measure of the quality of a screening test where a long lead time indicates an effective screening test, showing that screening can advance diagnosis effectively. However, lead time is intertwined with sensitivity, screening interval and overdiagnosis which makes the interpretation difficult. For PCs detected by PSA screening, the mean lead time has been reported to be up to 12 years.[19] However, this estimate is the mean lead time of all screen-detected tumors including the overdiagnosed ones. Overdiagnosed tumors have a preclinical detection phase that exceeds the remaining lifetime of the individual and their lead time is therefore infinity. These cases therefore inflate lead time estimations. Lead time for clinically significant tumors are probably much shorter, otherwise a screening interval of 10 years would be adequate, assuming a perfect screening situation. Studies have reported a wide range of mean lead times from 3-12 years. Draisma et al. suggested that there are at least three reasons for this: 1. the screening context i.e. the population at risk and the screening algorithm; 2. the definition of lead time and overdiagnosis used; 3. the method used for calculating the estimates.[155] Draisma et al. defined three different types of lead time: non-overdiagnosed lead time, censored lead time and uncensored lead time. They suggested that the type of lead time should be defined in future studies. Non-overdiagnosed lead time can only be calculated for non-overdiagnosed cancer (clinical diagnosis precedes death). Censored lead time can be calculated for both overdiagnosed and non-overdiagnosed cancers (lead time for overdiagnosed cancers are censored at time of death from other causes). Uncensored lead time can also be calculated for both overdiagnosed and non-overdiagnosed cancers but with the difference that the overdiagnosed lead time is not censored at the time of death from other causes. The conclusion from this work is that an estimated lead time applies exclusively to the situation from which it was derived.

Several studies have investigated non-overdiagnosed lead time using information obtained from stored serum samples, where the time from an elevated PSA to clinical PC diagnosis is measured. These studies have

reported lead time estimates ranging between 3 and 12.8 years.[30, 31, 156-158] One limitation with this type of study design is that it is assumed that the cancer would have been detected at the time of the raised PSA and that only cancers that would have surfaced clinically are detected (100 % sensitivity and 100% specificity). Another limitation is that the results are dependent on the length of follow-up. If the follow-up is not long enough, such that all cancers have had time to surface clinically, lead time will appear shorter than it really is.

Lead time can also be estimated as the prevalence:incidence ratio, where the calculation is based on the formula P=D*I (P=prevalence of preclinical detectable phase, D=mean sojourn time and I=incidence) and prevalence is estimated from detection rate (DR) at first screening and sensitivity (s) of the test (P=DR/s). [159-161] One limitation with the prevalence:incidence ratio is that it assumes that incidence can be treated as a constant which has not been the case for PC. Instead, the "catch-up time method" or incidence:risk ratio has been suggested as an alternative. With this method mean lead time is calculated as the time required for the cumulative incidence in the unscreened group to catch up with the detection rate in the first screening round.[161, 162] Applying this method to the five largest centers in the ERSPC, Finne et al. calculated the mean lead time in the ERSPC to 6.8 years.[162] Opportunistic screening in the control arm (contamination) can affect lead time so that it appears shorter than it is.

Lead time can also be estimated with computer simulation models. Modeling studies have several advantages; it is possible to control for e.g. contamination in the control arm, various secular trends and different screening strategies (by varying PSA thresholds, screening interval etc) can be investigates. For example, simulation models based on the Dutch section of the ERSPC estimated mean lead times and rates of overdiagnosis for screening at various ages and screening intervals. Mean lead time varied from 12.3 years for a single screen at age 55 to 6.0 years for a single screen at age 75.[19] Other studied have used US screening and PC incidence patterns to estimate mean lead times between 4.5 to 7 years [163, 164] Differences in these lead time estimates can partly be explained by different PSA-thresholds for biopsy and biopsy compliance rates.

As previously mentioned lead time and the rate of overdiagnosis are closely connected. Similar to lead time, overdiagnosis is difficult to calculate, however, it can theoretically be calculated as the absolute difference in the number of cancer cases between the study arms in a randomized controlled study where all study participants are followed for their life-time and there

has been no screening in the control group. Estimates of overdiagnosis in PSA-screening vary greatly across studies and range between 2.9-88.1%.[19, 155, 163-173] An article by Etzioni et al. discussed the reasons for this wide range of estimates and identified several different reasons described below. [174]

-Definitions and methods of measuring overdiangosis: From a study aiming at interpreting different overdiagnosis estimates in mammography screening de Gelder et al cited seven different ways of measuring overdiagnosis. Depending on which denominator was used and depending on which phase of screening was assessed the estimates could vary by a factor of 3.5 and 4, respectively.[175] Overdiagnosis has been reported as: 1) the fraction of screen-detected tumors, 2) the fraction of all tumors detected in the screening group and 3) the fraction of all men invited to screening. Other measures can be number of cases per adverted PC death (NND) or, number of overdiagnosed tumors relative to the number of cases expected without screening – for the entire population, or for the screening population. Overdiagnosis can also be expressed as the relative risk of cancer with screening compared with the relative risk of cancer without screening.[175]

-Screening context: the screening strategy has a large effect on the risk of overdiagnosis. The PSA threshold, the number of biopsy cores, screening interval, and age at screening will all affect the risk of overdiagnosis. A low PSA threshold and frequent screens will increase the risk of overdiagnosis. The incidence of PC in the comparison group, whether it is a control group or the expected PC incidence in the background population, will also affect the risk of overdiagnosis. For example, if there is widespread opportunistic screening in the comparison group, the overdiagnosis estimate in the screening group relative to the comparison group will be lower. If the background population has been opportunistically screened, the pool of latent cancers that can be detected by screening and the risk of overdiagnosis will be smaller.[174]

-Estimation approaches: “excess incidence approach” and “lead time approach” are the two main approaches for estimating overdiagnosis. The rationale behind the excess incidence approach lies in the observation of incidence patterns that follow an introduction of screening where the cancer incidence first increases and then declines to a new, higher level which indicates the amount of overdiagnosis. However, the estimates from this method are highly dependent on at which point in time the analysis is performed. If the introduction phase (“the prevalence round”) of a screening program is included, the estimates will be higher than if the estimates are

calculated when a screening program has reached "steady state". During the introduction phase of screening, when the prevalence screens are performed, the excess incidence will include both overdiagnosed and non-overdiagnosed cancers, compared to when the screening program has reached a steady state and the excess incidence only consists of overdiagnosed tumors. The age groups included in the analysis are also important. Screening should cause a "compensatory drop" in incidence in the older age groups, which have completed screening. If these age groups are not included in the analyses, overdiagnosis will be over-estimated. The "lead time approach" uses mathematical models to predict lead time and overdiagnosis based on the pattern of excess incidence. A third approach is to relate the excess incidence caused by screening to the cancer-specific mortality reduction, by calculating the NND. This estimate can give a picture of the harms-to-benefit balance of screening, but is strictly speaking not an estimate of overdiagnosis. Published estimates of NND have ranged between 5 and 48 (Table 5). Similar to the other excess incidence approaches, estimates of NND are highly time-dependent measures, which partly explain the wide range.

Published estimated of overdiagnosis is summarized in table 5. When reviewing these estimates, it becomes obvious that they are not directly comparable due to the reasons listed above. For example, many of the US estimates are lower than the European, which can be partly explained by higher PSA thresholds and lower compliance with biopsy following a positive screening test in the US. In addition, it is clear that PC screening with PSA results in substantial overdiagnosis, regardless of how it is estimated.

Author, year of publication	Study years	Population/context	Approach	Definition	Results
Draisma et al. 2003[19]	1994-2000	Dutch section ERSPC, MISCAN model, 55-74 years	Lead time	% of screen-detected cases	27-56% depending on age at screening and interval
Telesca et al. 2008[163]	1973-2000	US incidence trends, SEER data, 50-84 years	Lead time	% of screen-detected cases	22.7% for whites 34.4% for blacks
Draisma et al. 2009[155]	1985-2000	US population, SEER data, 54-80 years	Lead time	% of screen-detected cases in 3 different models	23-42% depending on model
Wu et al. 2012[165]	1996-2005	Finnish section of ERSPC, 55-67 years	Lead time	Number of screens for overdetection (NSO)	29
Zappa et al. 1998[166]	1992-1995	Italy, screening pilot study, 60-74 years	Excess incidence	Excess cancers/expected number of cancers in the absence or screening	51%
McGregor et al. 1998[168]	NR	Quebec Canada, annual screening 50-70 years	100-CFR (CFR= case fatality rate=rate screen-detected lethal cancers/rate screen-detected cancers	% of screendetected cases not causing death	84%
Etzioni et al. 2002[164]	1988-1998	Simulation model of men 60-84 years that best matched US incidence trends from SEER data	Excess incidence	% of men whose PC were screen-detected and who otherwise would not be detected	29% for whites 44% for blacks
Ciatto et al. 2005[169]	1991-1994	Italy, pilot screening program, 60-74 years	Excess incidence	Excess incidence observed/expected	66%
Pashayan et al.	1996-2002	Cambridge England, population	Excess incidence	% of PSA detected cases	40-64%

2006[167]		based, ages >40 years			
Pashayan et al. 2009[160]	2002-2005	ProtecT screening trial, English population based cancer register, 50-69 years	Excess incidence	% of PSA detected cases	10-31%
Welch et al. 2009[176]	1986-2005	US incidence , SEER data, ages ≥ 20 years	Excess incidence	Number of excess cases	1.3 million NND 23
Gulati et al. 2011[177]	Projected 25 year results	US population	Modeling/lead time	NNT/D	9
Loeb et al.2011 [178]	Projected 12 year results	ERSPC, European screening population	Projected excess incidence/mortality reduction	NND	18 at 12 years
Heijnsdijk et al. 2012[170]	Entire lifespan	ERSPC, European screening population, MISCAN model 1000 men of all ages, screening between 55-69 years	Modeling/lead time	NND and % of screen-detected cases	5 30-48% depending on age at screening and interval
Hugosson et al. 2010[2]	1995-2008	Göteborg screening trial, 50-64 years	Excess incidence/mortality reduction	NND	12 at 14 years
Schröder et al. [59, 148, 149]	1990's-2006, 1990's-2008, 1990's-2010	ERSPC, European screening population, 55-69 years	Excess incidence/mortality reduction	NND	48 at 9 years, 37 at 11 years, 27 at 13 years
Kilpelainen et al. 2013[179]	1996-2010	Finnish section of ERSPC, men 55-67 years	Excess incidence/mortality reduction	NND	25 at 12 years
Robool et al. 2013[180]	1993-2010	Rotterdam section of ERSPC, men 54-74 years	Excess incidence/mortality reduction	NND	33 at 12.8 years

Gulati et al. 2013[171]	Entire lifespan	US population, modeling 35 different screening strategies	Modeling/lead time	NND and lifetime probability of overdiangosis	NND 2.99-7.08 1.3-6.0%
Gulati et al. 2014[172]	1975-2005	US population, SEER data modeling, 50-84 years	Modeling data used to develop a nomogram	% of non-metastatic screen-detected cases	2.9-88.1% depending on age, Gleason score and PSA at diagnosis
de Carvalho et al. 2014[173]		US population, MISCAN, yearly screening 50-74 years	Modeling/lead time, 83 different screening polices	Lifetime risk of overdiagnosis	3.8%

Table 5. Estimates of overdiagnosis in different studies and populations.

As previously mentioned, overdiagnosis can also be viewed from a clinical perspective, trying to identify tumors that are unlikely to cause disease-specific morbidity or mortality even without treatment. Several attempts have been made trying to describe overdiagnosed cases, with synonyms such as "minimal", "focal", "insignificant" and "indolent disease". In many scientific papers, the definition of "indolent disease" is often based on strict pathological criteria, whereas the definition of "insignificant disease" also takes patient age and comorbidity into consideration. According to a review article, the most commonly used definition of "insignificant PC" is based on pathology at radical prostatectomy and includes: organ-confined disease (no extraprostatic extension, no seminal vesicle invasion and no lymphnode involvement), no Gleason pattern 4 or 5 and a maximum tumor volume of 0.5 mL.[181] The tumor volume threshold of <0.5 ml is based on a study by Stamey et al. with incidentally detected PCs in a cystoprostatectomy series from the 1990s. The authors hypothesized that the largest tumors were the most aggressive. Because the lifetime risk of clinically diagnosed PC was 8%, the authors chose the largest 8% of the tumors (destined to surface clinically during a man's lifetime) which all had a volume above 0.5 mL – hence, the cut-off 0.5 mL to define significant versus insignificant PC.[182] Another frequently cited paper is a series of 157 consecutive radical prostatectomies, where Epstein et al. defined "insignificant PC" as organ-confined, Gleason score <7 with a tumor volume <0.2 mL. The same authors also defined "minimal disease" as a tumor volume <0.5 mL and with the same criteria as for "insignificant disease".[183] However, it has been questioned whether this volume thresholds are applicable to screen-detected cancers. In addition, PC is oftentimes a multifocal disease, that is, a man can have not only one focus of PC, but several foci.[184] A more recent study based on the first screening round in the Rotterdam section of the ERSPC reported a total tumor volume threshold for the index tumor of 0.55 mL and a total tumor volume threshold of 0.7 mL. When accounting for tumor grade and stage these figures became higher, 1.3 and 2.5 mL respectively.[185]

A number of nomograms exist that help predict the risk of insignificant cancer. For men with a high probability of insignificant cancer, a conservative management such as active surveillance can be the first treatment strategy whereas men with a low probability (for example less than 30%) can be advised immediate curative treatment.[186] However, these probability cut-offs are arbitrary and should be set individually depending on patient characteristics and preferences. In the radical prostatectomy series by Epstein et al. discussed above, the model that best predicted insignificant PC included; PSA density (PSAD) <0.1 ng/mL, Gleason score <7, <3 cores with cancer, and ≤50% cancer in any core. These criteria could identify

insignificant cancer (organ-confined, no Gleason 4 or 5, and tumor volume <0.2 mL) with a PPV of 95%, NPV 66%, and an overall predictive value of 73%.[183] An updated version of the preoperative Epstein criteria was presented in 2004, based on a series of T1c tumors. The combination of PSAD<0.15 ng/mL, Gleason score ≤6, <3 cores with cancer and ≤50% cancer in any core correctly identified 91.6% as organ-confined at radical prostatectomy.[187] Current nomograms for predicting insignificant PC are based on findings at radical prostatectomy; this is, however, a surrogate endpoint for more long-term endpoints such as disease progression and disease-specific mortality. Three commonly used nomograms are the ones by Kattan, Steyerberg, and Nakanishi.[188-190] These nomograms use pretreatment clinical characteristics to predict insignificant disease at radical prostatectomy. For example, for a 60 year old man diagnosed with a T1c cancer, PSA 4 ng/mL, prostate volume 40 mL and 5 mm cancer in one biopsy core, these nomograms predict the probability of insignificant cancer ranging from 22-65% (lowest value Kattan and highest value Steyerberg).[191] These nomograms are based on different populations; some are based on patients referred to a urology clinic and others on the general population in a screening setting. This should be kept in mind when interpreting the different results and is also a reason for questioning the generalizability of these tools.

1.6.2 Why overdiagnosis a particular problem in prostate cancer screening

The lifetime risk for a Swedish man of being diagnosed with PC is approximately 16%, which can be compared to the lifetime risk of dying from PC for a Swedish man which is 5-6%.[192] This incidence:mortality ratio for PC is indicative on the presence of overdiagnosis. So why is overdiagnosis such a large problem for PC and PSA screening?

Tare several factors promoting overdiagnosis, and the most important are discussed here.

Firstly, of all, there is a large reservoir of latent, silent cancers that can potentially be detected with screening. As previously mentioned, autopsy studies of men who died from causes other than cancer have shown that PC can be detected as early as the 3rd decade of life and the prevalence of PC increases with age. Among men in their fifties and sixties, that is, men in the potential screening age, 30-60% harbor PC.[8, 9] Similar proportions of PC have also been reported from cysto-prostatectomy series where PC is found in up to 50% of the specimens.[193] Increased diagnostic activity with PSA, of both asymptomatic men and those with mild lower urinary tract symptoms

(LUTS), risks detecting these tumors which would otherwise have been diagnosed. The reported incidence of insignificant PC in radical prostatectomy specimens varies from 2.3-25% depending on the patients included.[181] In a screened population, Postma et al. reported minimal cancer (organ-confined, Gleason score ≤6 and tumor <0.5 mL) in 32-42% of radical prostatectomy specimens.[194]

Secondly, as already mentioned, PC incidence increases with age. Many parts of the world have an ageing population and men's life expectancy is increasing. For example, Swedish 65 year-olds have a remaining life expectancy of more than 18 years.[195] Similar estimates can be seen in other parts of the Western world.[196] The benefits of early diagnosis are probably smaller for older men compared to younger.[197] Increasing age is associated with an increased number of comorbidities and the risk of dying *with* PC rather than *from* it increases. Several studies have reported that age is an important risk factor for overdiagnosis of PC.[172, 173, 198]

Thirdly, PSA is highly specific for prostatic tissue but it is not cancer-specific. PSA as a screening tool have several limitations, see paragraph 1.5.1. Inflammation/infection, BPH and trauma can cause PSA elevations and poorly differentiated, highly lethal, cancers sometimes do not even produce PSA and are therefore no accompanied by increased PSA-levels which makes detection difficult. In addition PSA cannot differentiate between low- and high-risk cancers. The limited specificity implies that many men will have to undergo biopsies with the risk of detecting insignificant cancer that have no relation to the PSA elevation. In addition, screening is associated with length time bias and risks detecting a dis-proportional amount of less aggressive, slow-growing tumors, which may not pose a threat to life, while missing the aggressive tumors for which screening was intended.

1.6.3 Consequences of overdiagnosis

Overdiagnosis has consequences on all levels; for the individual patient as well as for society as a whole. The psychological consequences of overdiagnosis are obvious; "healthy" men, who in the absence of screening would never have been aware of their PC, are turned into patients. More men will have to go through work-up following a positive screening test including the discomfort and risk of complications of prostate biopsy. Thereafter follows a period of waiting for the biopsy results and a substantial portion of these men will subsequently receive a cancer diagnosis. The harms of PC screening, which would affect more men if screening were to be introduced, due to overdiagnosis, are:

- *PSA blood test*. From the PLCO it was reported that complications following a PSA blood test was seen at a rate of 26 per 10 000 screens. There were only minor complications reported such as dizziness and, bruising and hematoma from the blood draw.[146]

- *Biopsy*. Common complications following prostate biopsy include hematospermia, hematuria, hematochezia, fever, urinary tract infections and urine retention. Bleeding complications (hemostospermia, hematuria, hematochezia) are often mild and self-limiting. Macroscopic hematuria is common with reported rates of up to 84%, but hematuria requiring hospitalization occurs in less than 1%.[199] From the Dutch arm of the ERSPC hematuria lasting longer than 3 days, hematospermia and rectal bleeding were reported after 22.6%, 50.4% and 1.3% of the procedures, respectively. Other series have reported rectal bleeding in 30-40%. Although massive rectal bleedings are uncommon they can be life-threatening and require intervention.[200] Despite local anesthetics, some men find the procedure painful and unpleasant but the majority rate these symptoms as a minor problem and would be willing to undergo a repeat biopsy, if necessary.[200, 201] Infectious complications including urinary tract infections, epididymitis, fever and sepsis have increased, most likely due to microbial antibiotic resistance.[202, 203] Most infectious complications are caused by the gram negative bacteria Escherichia Coli.[204] Infectious complications requiring hospitalization have also increased and up to 6.3% of men undergoing a biopsy require hospital admission within 3 days, most for febrile infections.[199, 205] A recent report from the Prostate Cancer Data Base (PCBaSe), Sweden showed that infectious complications are increasing also in Sweden and 1% of biopsied men were hospitalized for a post-biopsy infection. The strongest risk factors for an infections complication not requiring hospitalization were a history of UTI and severe comorbidity. Infectious complications requiring hospitalization increased with time (from 2006 to 2011) and were more likely in men with a high comorbidity score.[203] Acute urinary retention is uncommon (<2%) but worsening LUTS are reported in <25% of patients.[199]

Recently there has been a growing concern that prostate biopsies, especially repeated biopsies and extended biopsy schemes may cause erectile dysfunction. The hypothesis is that biopsies may cause inflammation that may temporarily damage the neurovascular bundles because of their anatomical proximity to the prostate, and/or that biopsy may hit the nerves and cause direct damage. Psychological effects of undergoing work-up for cancer may also play a role. The scientific evidence is sparse but there seems to be a trend towards increasing erectile dysfunction during the first month but that these changes resolve with longer follow-up.[199] The long-term effects of repeated biopsies for men on active surveillance need further investigation. The risk of a fatal complication following prostate biopsy is extremely low and men selected for prostate biopsy are healthier than the general population.[203, 206, 207]

- *Anxiety*. A review by Hewitson concluded that health-related quality-of-life (HRQoL), psychological distress and anxiety were not dramatically increased for asymptomatic men during the screening process.[208] Carlsson et al. investigated the level of anxiety in men with a positive screening test (PSA ≥3 ng/mL) in the Göteborg screening study and showed that levels of anxiety were generally low; 66% reported no anxiety while awaiting the PSA result and 45% reported no anxiety associated with prostate biopsies. There was no significant association between the PSA-level and anxiety levels. However, for a small subgroup of patients, screening was associated with severe anxiety.[209] Several other studies have reported similar results.[210, 211] Wade et al. recently reported from the ProctecT-trial that levels of anxiety and depression associated with prostate biopsy were low and comparable with those of the general male population with the exception of two groups of patients; those with problematic post-biopsy side-effects and those who received a cancer diagnosis.[212]

  Many patients receive the receipt of a cancer diagnosis as a death sentence. The psychological consequences of living with a cancer diagnosis have been studied in both qualitative and quantitative studies, albeit the psychological harms of overdiagnosis are less well studied than the physical

consequences. Several studies have reported increased rates of anxiety, depression, post-traumatic stress disorder, cardiovascular events and even suicide for men living with a PC diagnosis.[213-216] Physicians need to be aware of these potential consequences of living with a cancer diagnosis, in order to offer appropriate counseling and/or pharmacological treatment of anxiety and depression, when indicated.

- *False positive results*. Men with false positive PSA results have more PC specific worry both in the short and long term than men with true negative results.[217] Men with false positive results also have more subsequent PSA testing and visits to a urologist.[218] False positive tests also lead to unnecessary biopsies. It is therefore important to minimize this rate. The rate of false positive tests depends on the PSA cut-off and the number and location of biopsies. In the Göteborg screening trial, using a PSA cut-off ≥3ng/mL, the PPV was 24% meaning that 76% of those with a positive screening test had negative biopsies.[119]

- *False negative result*. False negative results occur due to a false negative PSA results, i.e. a PSA below threshold despite cancer in the prostate, or due to a false negative biopsy, i.e. a PSA above threshold and PC present in the prostate but PC is not detected in the biopsy. The rate of false negative PSA results is dependent on PSA threshold and the rate of false negative biopsies depend on the biopsy technique and the number of sampled biopsy cores. For sextant biopsies, false negatives range between 15-34%. [133]. A false negative screen can result in delayed diagnosis and in the worst case, a missed chance of cure.

- *The economic consequences*. The economic consequences of overdiagnosis have been very little studied. The costs of introducing a screening program, as performed in the Dutch section of the ERSPC (age 55-75, PSA threshold 3 ng/ml, sextant biopsy, screening interval 4 yrs), have been estimated with the MISCAN model. Compared to a situation with no screening at all, health care costs for PC would be doubled if screening (100% attendance rate) was implemented. Only 10% of the additional costs would be due to the screening program itself whereas diagnosis and

treatment of overdiagnosed men would account for almost 40% of the total cost. Screening with a 4 year interval up to the age of 75 was the most expensive screening strategy, both annual and biennial screening up to 70 years were less expensive.[219]

1.6.4 Overtreatment and quality-of-life issues

In addition to increasing health-care costs and the many potentially negative psychological consequences of living with a cancer diagnosis, overdiagnosis may lead to subsequent overtreatment, which is believed to be the major problem with overdiagnosis. Many patients receive treatment and end up being treated for a cancer that would never have surfaced clinically, which is synonymous to overtreatment. This would be less of a problem if the treatments did not have side-effects. However in the case of PC, treatments with curative intent (radical prostatectomy and radiotherapy) have long-term side-effects with consequences for QoL. In addition, as screening advances diagnosis, men have to live longer with these side-effects. The largest reduction in HRQoL occurs within the first 1-2 years after treatment and stabilizes thereafter. Differences in the effects on various HRQoL domains between treatment alternatives tend to attenuate over time.[220] Common long-term side-effects following curative treatment for PC are urinary incontinence, impotence and bowel disturbances. Radical prostatectomy has a greater effect on sexual and urinary function while radiation therapy had a larger effect on bowel function, at least during the first 5 years.[221, 222] Reported rates of postoperative incontinence (0-65%) and impotence (29-100%) vary greatly in different reports.[223] Baseline function, patient age and operative skills of the surgeon all have an important influence on the functional (and oncological) outcomes. Historically, erectile dysfunction used to occur in almost all patients post radical prostatectomy but with the advent of the nerve-sparing technique [224] some men are potent postoperatively. In the SPCG-4 trial, which compared radical prostatectomy to watchful waiting, 84% reported erectile dysfunction and 41% urinary leakage post-operatively after a median follow up of 12 years, compared to 80% and 11% for watchful waiting patients.[225] Urinary incontinence is at its worst the first moths postoperatively and then improves. Carlsson et el. recently published a paper on the effect of surgical variability on the functional and oncological outcome of men undergoing open radical prostatectomy at the Sahlgrenska University Hospital in Göteborg, Sweden. Of men reported to be potent at baseline, 19% reported being potent 18 months postoperatively, and of those reporting to be continent at baseline 85% were continent 18 months postoperatively. There

was a large heterogeneity between surgeons with respect to urinary continence, with incontinence rates ranging between 7% and 30%, depending on which surgeon performed the operation.[75] To this date, no RCT has demonstrated any difference in functional and oncological outcomes between open radical prostatectomy and robot-assisted radical prostatectomy.[11, 64] In a study from the Göteborg screening trial it was shown that if screening were to be implemented, the "cost" of saving one man from PC death is 4 men impotent and less than one man incontinent.[226] The 30-day mortality after radical prostatectomy is low at 0.1-0.5%.[227]

During the last decade there have been significant improvements in the planning and delivery of radiotherapy, which has reduced complication rates. Complications are usually divided in acute and late toxicity (≥ 6 months post treatment). Acute toxicities occur in about 50% of patients and include LUTS (mainly irritative symptoms), hematuria, diarrhea, rectal bleeding or mucus discharge and anal irritation. These are usually self-limiting and respond well to symptomatic treatment. Late toxicities are also mainly gastrointestinal and genitourinary symptoms and include increased bowel frequency, rectal bleeding, urethral strictures, LUTS, hematuria (ranging from minor hematuria to severe hemorrhagic cystitis), erectile dysfunction and secondary malignancies such as rectal cancer. The exact mechanism behind erectile dysfunction after radiotherapy is unclear but might be due to vascular damage and scarring. Acute and late complications are often classified according to the Radiation Therapy Oncology Group (RTOG)/European Organization for Research and Treatment of Cancer (EORTC) morbidity scale which is graded from 1 to 4 (1=mildest, 4=fatal). In an EORTC trial, late toxicity after external radiation (70 Gy) was reported for 377 men with T3-4 tumors. A total of 22.8% had grade ≥2 urinary or intestinal complications or edema in the lower extremities (legs). Grade 3 or 4 complications occurred in <5% of patient but 4 of the 377 men died due to grade 4 complications.[228]

Side effects from treatment also have consequences for various aspects of QoL. A large prospective study by Sanda et al. investigated factors that were associated with changes in QoL radical prostatectomy and radiotherapy reported by patients and their partners. Each treatment alternative resulted in a distinct pattern of changes on QoL related to the various side-effects. The degree of satisfaction, for patients and their partners, regarding treatment outcomes was significantly associated with changes in QoL domains affected.[229]

A recent paper by Schaake et al. investigated the impact of radiation-induced toxicity on HRQoL. They found that also mild, grade 1 toxicity, had negative

effects on HRQoL. Urinary incontinence affected all HRQoL scales and the effect was most pronounced 6 months after therapy, but the negative impact decreased thereafter. Rectal discomfort was common (18%) and had a major impact on HRQoL. The proportion of men who experience any rectal discomfort was stable over time, but the level of discomfort decreased. There was no significant association between HRQoL and radiotherapy technique, adjuvant hormonal treatment or patient characteristics.[230]

From the PLCO trial, Davis et al. reported on long-term side-effects of PC treatment and effects on cancer-related and general QoL. Urinary, sexual, and bowel side effects were independently associated with cancer-related QoL. Urinary and bowel side-effects were also associated with general QoL. Bowel side effects had the strongest association with all QoL outcomes. Treatment related side-effects were present and affected QoL up to 10 years post-treatment.[231]

With data from the ERSPC trial Heijnsdijk et al. used the MISCAN model to predict quality adjusted life-years (QALYs) gained after introducing organized PSA-screening compared to a situation without screening. Over the lifetime of 1000 men, who would be screened annually between 55-69 years, there would be nine fewer PC deaths and 73 life years gained. However, this figure would have to be adjusted by -23% due to loss in QoL, resulting in 56 QUALYs gained. The main reasons for the loss in QoL were overdiagnosis and side-effects from curative treatment. The model also predicted that over the life-time of the 1000 men, 45 men would be overdiagnosed and overtreated and there would be a loss of 1134 PC-free years (lead time years). [170]

Active surveillance has emerged as a treatment strategy with the aim of reducing the overtreatment of screen-detected PCs as described in paragraph 1.2.4. The challenges of active surveillance are discussed in depth in the discussion section of paper II.

1.7 Screening and overdiagnosis in other fields of medicine

Overdiagnosis and overtreatment are not unique for prostate cancer screening. Both phenomena are common in many field of modern medicine. Not all individuals who receive an intervention, whether it is a screening test, a medication or an operative procedure will benefit from it. A certain level of overdiagnosis and/or overtreatment has to be accepted. For example, when

treating mild to moderate hypertension with thiazide diuretics, 122 patients need to be treated during 5 years to prevent one cardiovascular event.[232]

Mammography

The benefits and harms of breast cancer screening with mammography have been debated extensively during the last decades. Similar to any cancer screening, the benefit of a reduction in disease-specific mortality has to be weighed against a number of potential negative consequences. Some women find the examination painful and unpleasant. The psychological effects of a false-positive examination are well-studied.[233] False positives occur at a rather high rate in breast cancer screening. According to a Cochrane review; 200 women will receive a false-positive result were further investigations will show that they do not have cancer if 2000 women are screened during a 10 year period.[234]

Analogous with PC screening, there is potential for overdiagnosis with mammography. There is a "disease reservoir" of asymptomatic cancers that can be detected with screening but that might never have surfaced clinically. Autopsy studies have shown that undetected breast cancer lesions are common and found in 7-39% of middle-aged women.[235] Since the 1960's a number of RCTs of mammography have been published, several of which have been conducted in Sweden.[234]

The USPSTF updated their recommendations for breast cancer screening in 2009. For women age 50-74 they recommend biennial screening with mammography. Their decision was based on a meta-analysis of six trials among women age 50-59 years that showed a RR for breast cancer death of 0.86 (CI 0.75-0.99) and NNI 1339 (CI 322-7455) and two trials among women aged 60-69 years that gave a pooled RR of 0.68 (CI 0.54-0.87) and NNI 377 (CI 230-1050). For women for women 39-49 of age they concluded that the net benefit was small (NNI of 1904) and therefore recommended against screening. They found only one study that reported results for women older than 70 years. This study could not show a reduction in breast cancer mortality.[236, 237] A 2011 Cochrane review by Gøtzsche et al. was more pessimistic about mammography and concluded that screening likely reduces breast cancer mortality by 15% but with approximately 30% overdiagnosis and overtreatment and that it is unclear whether breast cancer screening does more good than harm.[234] An independent expert panel in the UK reviewed the benefits and harms of breast cancer screening and published their conclusions in the *Lancet* in 2012. They estimated that screening with mammography reduces breast cancer mortality by 20% (RR 0.80, CI 0.73-0.89) and that approximately 19% of the cancers detected in the screening

group during the screening years were overdiagnosed. For women invited for screening during a period of 20 years starting at the age of 50 years NNI was 235.[238]

When Sweden introduced population-based mammography screening in 1986 it was as one of the first countries in the world. Since 1997, mammography is offered nationwide in Sweden. The National Board of Health and Welfare recommends screening for breast cancer with mammography every 18-24 months in women aged 40-74.[239] Critical voices have been raised that the invitations for mammography in Sweden are not detailed enough and should better state the advantages and disadvantages of screening, and that many women do not perceive the invitation as voluntary but rather as something mandatory.[240] The National Board of Health and Welfare is currently working on a revision of the invitation letter.

In summary, breast cancer screening with mammography reduces breast cancer mortality by 15-25%. The amount of overdiagnosis is difficult to estimate but ranges between 11-30% depending on method of calculation and the composition of the underlying population in which it is being studied. NNI varies between different reports from 100 to 2000 women, depending on age, duration of screening and length of follow-up.[238]

Colorectal cancer screening

The majority of colorectal cancer tumors develop from precancerous adenomas to carcinomas, a process that takes several years. Screening for colorectal cancer aims at detecting these precancerous lesions with fecal occult blood-testing (FOBT) and/or endoscopy (flexible sigmoidoscopy or colonoscopy). FOBT can be tested with the guaiac-method or with a human-specific immunochemical test.[241]

Four RCTs have investigated whether screening with FOBT followed by endoscopy reduces colorectal cancer mortality. [242-245] A Cochrane review combining the results of these four trials reported a reduction of 16% in colorectal cancer mortality (RR 0.84, CI 0.78-0.90) after 11-18 years of follow-up. There was no significant difference in all-cause mortality. [246] NNI for biennial screening with FOBT range between 600-1200 with up to 17 years of follow-up.[247]

Flexible sigmoidoscopy has also been shown to reduce disease-specific mortality, in five separate RCTs by 22-31%.[248-252] As colorectal cancer screening reduces the incidence of invasive cancer through detection and

removal of adenomas overdiagnosis cannot be estimated as NNT/NND.[249, 250, 252]

Negative aspects of colorectal cancer screening includes false negative tests (polyps that do not bleed or that do not bleed at the time of the test), false positive FOBTs leading to unnecessary endoscopies, overdiagnosis of polyps that would never have developed to invasive cancer and complications of endoscopic procedures (perforation, bleeding).[246]

In Sweden there is an ongoing screening program for colorectal cancer with FOBT in the Stockholm/Gotland region since 2008, and since 2013 the Swedish National Board of Health and Welfare recommends screening for colorectal cancer with FOBT for men and women aged 60-74. An almost nation-wide screening study is launched in 2014 comparing screening with FOBT and endoscopy.[253]

Cervical cancer screening

Screening for cervical cancer with Papanicolou (Pap) smear has a long history. Already in 1928 Papanicolou published a report indicating that cervical cancer could be diagnosed from exfoliated cells. Exposure to certain strains of human papilloma virus (HPV) has an important role in the development of cervical cancer and the majority of cases are preceded by a history of HPV-infection resulting in cellular abnormalities. However, most episodes of HPV infection are transient and would never develop into high-risk lesions or cancer. Screening for cervical cancer with Pap-smear and/or HPV-testing aims at diagnosing precancerous lesions and preventing them for develop into invasive cancer and thereby reducing cervical cancer incidence and mortality. The effectiveness of Pap smear screening has never been demonstrated in a RCT. Evidence comes from observational studies and by studying trends in cervical cancer incidence and mortality following the introduction of screening, which has shown that organized screening with Pap smear can reduce cervical cancer mortality by up to 80%.[254-257] The Nordic countries introduced organized screening at different times; Finland was first in 1963, soon followed by Iceland and Sweden. Denmark also launched a program already in 1964 but has had problems reaching an adequate coverage in the population; the attendance rate was only 45% as late as in 1991. In Norway, screening was opportunistic until 1994 when an organized screening program was introduced. As expected Iceland, Finland and Sweden have had the most pronounced incidence and mortality reductions, 55-75% and 60-76%, respectively, while the corresponding figures for Denmark and Norway have been lower.[256] NNS to prevent one

cervical cancer death has been estimated to 1140 after 10 years for regular Pap smears.[258]

Similar to other screening method, there are potential harms with screening and treatment for cervical cancer. Abnormal tests can lead to more frequent testing and invasive procedures (colposcopy, cervical biopsy) resulting in bleeding, pain and infections.[259] An abnormal test can also lead to a short-term increase in anxiety and distress. The potential harms of treatment include adverse pregnancy outcomes such as preterm delivery.[260] There is also a risk of overdiagnosis. Many precancerous lesions will never develop to invasive cancer but regress and other lesions are slow-growing and will never become symptomatic during the woman's lifetime. Similar to colorectal cancer screening, cervical cancer screening aims at reducing the incidence of cervical cancer screening by removal of precursors, therefore NNT/NND cannot easily be estimated.

The USPSTF recommends screening for cervical cancer with cytology (Pap smear) every three years for women aged 21-65, alternatively screening every 5 years with a combination of cytology and HPV-testing for women aged 30-65.[261] The Swedish national screening program for cervical cancer is currently reviewed. Today, women are regularly invited (every 3-5 years) for a Pap smear between the ages of 23 to 60. In some parts of the country, testing for HPV is also part of the screening program. As a complement to the national screening program Swedish girls (10-12 years) are offered vaccination against HPV strains number 16 and 18, which are most strongly associated with increased risk of cervical cancer, as part of the vaccination program for children in Sweden.

2 AIM

The overall aim of this thesis was to explore different aspects of overdiagnosis in screening for PC with PSA. The specific aims of each paper were as follows:

Paper I

To investigate the accuracy of Swedish COD certificates for men with PC to ensure that these can be used for endpoint evaluation in screening studies for PC and to investigate whether overdiagnosis affected accuracy.

Paper II

To evaluate whether active surveillance can be used as a treatment strategy for screen-detected PC to reduce overtreatment.

Paper III

To investigate the effectiveness of organized compared to opportunistic screening in reducing PC mortality, measured as number needed to invite, as well as risk of overdiagnosis, measured as number needed to diagnose

Paper IV

To investigate the effect of age and number of screening visits on the risk of being diagnosed with PC.

3 PATIENTS AND METHODS

3.1 Population

The Göteborg randomized population-based PC screening study is the base for all four papers (I-IV) in this thesis. This trial was initiated in 1994 following approval from the ethical review committee at the University of Göteborg. As of Dec 31 1994, there were 32,298 men born between 1930 and 1944 (age 50-64, median 56 years) and living in the city of Göteborg, according to the Population Registry. Using computer randomization, 10,000 men were randomized to a screening group and 10,000 men to a control group. A total of 101 men were excluded (50 in the screening group and 51 in the control group); 56 men had prevalent PC, 34 had deceased and 10 had emigrated but had not been removed from the Population Register at the time of randomization; another man in the control group refused participation (Figure 7). No informed consent was deemed necessary for men in the control group. Men in the screening group received a letter with information regarding PSA and its advantages and disadvantages together with an invitation for PSA-testing every second year. The mean upper age limit for invitation was 69 years (67-71). The PSA threshold that led to a recommendation for further urological work-up was initially 3.4 ng/mL but has been lowered on two occasions for consistency with other ERSPC sites and due to a change in assay-calibrator (in 1999 to 2.9 ng/mL and in 2005 to 2.5 ng/mL). The urological work-up consisted of a DRE, TRUS and prostate biopsies. Laterally directed sextant biopsies were the standard up to year 2009, thereafter a ten-core biopsy scheme has been used. Men with a PSA below the threshold and men with benign biopsy were re-invited for PSA-testing after 2 years. The treatment strategy following a PC diagnosis was not specified in the study protocol, but was at the discretion of the patient and the treating physician. The vast majority of men in both the screening and control arm has been followed and treated by the same team of urologists at Sahlgrenska University hospital in Göteborg, Sweden.

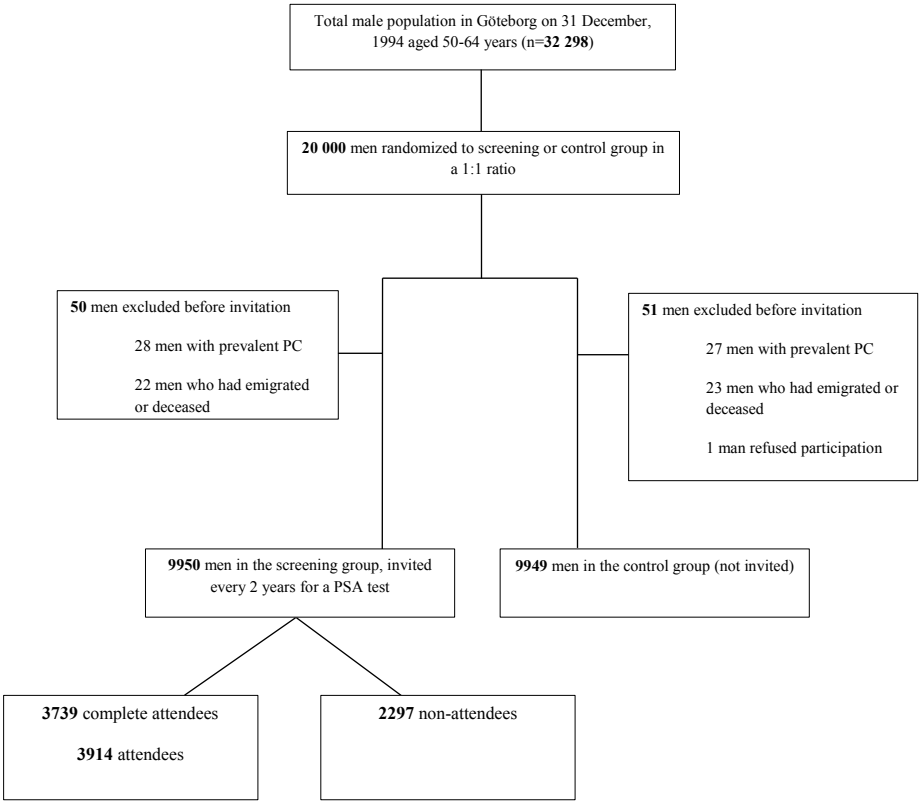

Figure 7. Consolidated Standards Of Reporting Trials (CONSORT) diagram showing the screening algorithm of the Göteborg randomized population-based prostate cancer screening study 1995-Jun 30, 2014

In all men with a PC diagnosis, both in the screening and control group, all available medical information regarding tumor stage, treatment and disease course was continuously entered into a study database. This database was linked every third month to the Regional Cancer Register (since 2009 also to the Swedish National Cancer Register, SCR) and the Swedish Population Register (SPR) to obtain information on PC diagnosis, mortality and emigration outside Sweden. In addition, for all deceased men, a copy of the COD certificate was obtained. COD in all men with a PC diagnosis was determined by an independent COD committee as described below.

As of spring 2014 the 10th and final screening round was completed when all age cohorts had reached the upper age-limit for invitation. However, the study data-base continues to be regularly linked to the SCR and the SPR to retrieve information regarding PC diagnosis and deaths.

3.2 Registers

Several different Swedish registers have served as important sources of information for the Göteborg screening study and for the four papers in the present thesis. Nordic countries, Sweden included, are known to have a high quality of their official registers. The Swedish 10-digit personal identity number, which is a unique personal identifier, is given to all individuals who are registered in Sweden. It is an important tool for linking medical registers and enables almost 100% coverage in many Swedish registers.[262]

Cause of death register - Dödsorsaksregistret

Sweden has a long history of population registers. Way back in 1749 a population register, "Tabellverket", covering the entire country, was established. At that time, the clergy was responsible for the reports. COD determination for all causes of death has been mandatory in Sweden since 1911, and since then, the National Board of Health and Welfare publishes a statistical summary report entitled "Causes of death" every year. Previously, cause of death was recorded in churches' funeral books and summaries were sent monthly to the district medical officer who reported to Statistics' Sweden. In 1991 the certificate was divided in two parts; one death certificate (*sv.* "Dödsbevis") which is sent to the Swedish Tax Agency as well as the Population Register, and another COD certificate (*sv.* "Dödsorsaksintyg") which is sent to the National Board of Health and Welfare that constitutes the base for COD statistics. Both certificates should be issued by a medical doctor.[263]

The COD certificate has two parts. Part I has four lines (A-D) and the bottom line (lined) should state the underlying COD and any other conditions entered on line A-C should be a direct consequence of the underlying COD stated on D. The underlying COD is the condition that started the chain of events leading to death, for example: PC → deep venous thrombosis → pulmonary embolism. Part II should state any other conditions that might have contributed to death, but which are not related to or a direct consequence to the COD, e. g. hypertension. Incomplete or inconclusive certificates are returned to the certifier for supplementation (2.7% of all certificates in 2008). If no COD certificate is issued, a reminder is sent to the doctor who issued the death certificate. The COD certificate was not issued in 1.84% of all deaths, despite reminders. In these instances the COD is coded as R99.9. Trained coders at the National Board of Health and Welfare code all deaths according the International Classification of Diseases - 10 (ICD-10) and enter the information in the COD registry. The reported underlying COD is changed from what is stated on the certificate in 20% of all cases.[264]

The autopsy frequency is the proportion of deaths where the COD is based on an autopsy (clinical or forensic). Autopsy is believed to be the most thorough examination to confirm a COD; declining autopsy rates over time may lead to more inaccurate COD determinations in national statistics. Since 1970, the autopsy frequency in Sweden has declined from 50 % to 12% in 2011. The autopsy frequency was 7% for women and 15% for men in 2011.[265]

The Swedish Cancer Register - Cancerregistret

The SCR was started in 1958 and is, since the 1980's, divided into six regional cancer registers which are managed by the Regional Cancer Centers (RCC). The RCCs are responsible for registration, coding, data check and correction work. Every health care provider is obliged, by Swedish law, to report every newly diagnosed cancer to the RCC, regardless of the mode of diagnosis (clinical, morphological, based on laboratory test or on autopsy).[266] The SCR is generally considered to be of high quality with 99% of all cancers morphologically verified [266] and an almost complete coverage. The completeness of the SCR was investigated in a sample survey in 1998. Possible underreporting was investigated by comparing the Hospital Discharge Resister with the SCR. The degree of underreporting was site-specific and increased with age. Underreporting for urological cancers was low. In total, underreporting was estimated at 3.7% of all reported cases.[267]

The Swedish Population Register -Folkbokföringsregistret

The Tax Agency (*sv*. "Skatteverket") is responsible for the SPR. The register contains information about people who live in Sweden and where in the country they reside. The first time a person is registered in the Population register he or she receives a personal identification number. The register receives information regarding migration within the country and emigration to other countries. When a person has deceased, the register receives a death certificates.

Paper I

This paper evaluated the quality of Swedish COD certificates in men with PC, to ensure that death certificates can be used for endpoint evaluation in a PC screening study. As PC mortality is the main endpoint in the Göteborg screening trial, an accurate COD determination is crucial. All men with a PC diagnosis (PC as underlying COD on the death certificate or identified through linkage with the SCR) and who deceased during the study period (Jan 1, 1995 - Dec 31, 2008) were included. Men lacking an analyzable death certificate were excluded (n=7, 2.5%). Swedish COD certificates have, as previously described, two parts. In this analysis, PC was only regarded as the

underlying COD if it was mentioned on the first part (line A-D) of the certificate.

All available medical information was retrieved for men in the study. This included hospital charts, reports from outpatient clinics, x-rays, histopathology reports and autopsy protocols. A COD committee, consisting of three experienced urologists, reviewed all collected material and decided on the COD using a pre-specified algorithm. This algorithm has been used in the whole of the ERSPC for cause of COD determination.[268] It consists of a hierarchy of questions that need to be answered by the committee members. The first set of questions is:

1. Were clinical metastases present and if so, were these PC metastases? The answer should be based on: clinical picture, PSA-level, x-ray, scans, treatment or no treatment, and pathology. Were there signs of progressive disease? The answer should be based on the same parameters as stated above. Was metastatic, progressive PC the COD? If in doubt or negative the next question is:
2. When in doubt or negative, the next set of questions is: Was there a progressive clinical recurrence? Was the progressive local recurrence the COD?
3. When in doubt or negative the next set of questions is: Where complications of primary treatment the COD? If negative, were complications of screening or biopsy the COD?
4. When still in doubt or negative, the last question is: Was the specific COD known and did PC contribute to this other COD?

The algorithm can lead to eight different endpoints; definitely PC death, probable PC death, definitely intervention-related death, possible PC death, unlikely PC death, PC as a contributory factor to death, definitely not PC death and pending, if information for deciding on COD were lacking.

In paper I, the COD was regarded as a dichotomous variable where definitely PC-, probable PC-, and intervention-related deaths were classified as “PC death” and all remaining endpoints as “other CODs”. Each committee member followed the algorithm and independently reached one endpoint. In cases of disagreement, consensus was reached by discussion in most cases but if no consensus could be reached the COD was a majority decision. Cases of special interest could also be discussed at a meeting with COD committees

from the other ERSPC centers. The committee was blinded for information regarding study arm in the majority of cases and only had access to the COD certificates in extraordinary cases where there was no other medical information available.

Sensitivity was calculated as the proportion of deaths correctly coded as PC deaths and specificity as the proportion of correctly coded non-PC death, with the committee's decision as reference.

Agreement between members in the COD committee was classified as total agreement, minor disagreement or major disagreement. Total agreement was when all three committee members had classified COD as the exact same endpoint, minor disagreement was when one or two of the members had classified different endpoints but where the final endpoint was the same, for instance if two members had classified COD as definitely PC and the third member had classified COD as probable PC (all classified as PC death in the final evaluation). Major disagreement was when committee members did not agree on the COD in the final evaluation. For example, two members had classified COD as definitely PC and one member as unlikely PC (final evaluation 2 PC deaths and 1 non-PC death).

To investigate factors that potentially could affect COD determination, all medical records were manually reviewed by the author of this thesis. We choose to register the presence of other malignancies (at the time of death or in the medical history), serious comorbidity, place of death and whether or not an autopsy had been performed.

After paper I was published in 2011, the COD committee has finalized the evaluation of COD for all men with PC up to Dec 31, 2012. Therefore, results up to this date are also presented in this thesis.

Paper II

This second paper assessed outcomes for men managed with active surveillance. All men who had active surveillance as the primary treatment strategy between Jan 1, 1995 and Dec 31, 2010 in the Göteborg screening study were included. The treatment strategy was not determined by the study protocol, but at the discretion of the patient and his treating physician. The reason for choosing active surveillance was in the majority of cases, a presumed low-risk cancer but could also be due to comorbidities or patient preference. There was no protocol determining the exact interval between follow-ups or indications for re-biopsy but follow-up was determined by the biopsy outcome, disease activity, patient's age and degree of comorbidity. All

men included in this study were managed by a small group of urologists, working under the same treatment policies, at a single university clinic. Typically, men were followed every 3-12 months with PSA and clinical examination, but the interval between follow-ups depended on disease activity and age. Most men were followed with active surveillance (curative intent) but as men grew older and comorbidities surfaced, active surveillance could gradually transcend into watchful waiting (symptomatic treatment). An early re-biopsy was performed if the initial biopsy contained a very small volume of cancer (<2 mm). Re-biopsy was also performed if there were signs of disease progression or typically, every two to three years in men with stable disease. Patients were recommended to remain on active surveillance as long as they felt comfortable with this management strategy, and as long as there were no signs of disease progression. Progression of PSA, grade or stage or led to a recommendation for a switch in treatment strategy to radical prostatectomy, radiation therapy or hormonal treatment. The patient himself could also request a change in treatment strategy. Follow-up continued every 3-6 months after treatment.

Tumors were divided in to risk groups (Table 6); the risk group definitions were a modified version of the D'Amico criteria.[55]A very low-risk group was also defined, based on the Epstein criteria for insignificant cancer.[183]

	PSA	Gleason score	T stage	Biopsy result
Very low	PSA density <0.15 ng/mL	≤6	T1c	<3 cores with cancer and ≤50% cancer in any core
Low	<10 ng/mL	≤6	T1	-
Intermediate	<20 ng/mL	≤7	T1-2	-
High	<100 ng/mL	≥8	T1-4	-
Advanced	≥100 ng/mL	-	M1 and/or N1	-

Table 6. Risk group definitions

Endpoints were overall survival, PC-specific survival, treatment-free and failure-free survival. Failure after active surveillance was defined as death from PC, development of PC metastases, initiation of hormonal therapy or PSA recurrence following surgery or radiation. PSA recurrence was defined as PSA ≥0.2 ng/mL after radical prostatectomy with or without salvage radiation and as the PSA nadir + 2 ng/mL after radiation therapy.

Paper III

This paper aimed at investigating the effectiveness of organized versus opportunistic screening in reducing PC mortality, measured as NNI, and the amount of overdiagnosis, measured as NND. The screening group in the Göteborg screening study has, as previously mentioned, been invited to organized PSA screening every second year since 1995. In contrast, men in the control group have not been invited to such screening and have been unaware of their participation in a screening trial for PC. However, since the study started in 1995 opportunistic PSA screening has increased dramatically in Sweden.[269, 270] This implies that men in the control arm have also been exposed to these increasing levels of opportunistic screening. In the study data-base, information regarding the method of detection was prospectively registered for both the screening and control group.

In 1990-94 PSA-usage as a screening test was close to non-existing in Sweden. Based on PC incidence data from the SCR and mortality data from the SPR for this pre-PSA-era years, we were able to estimate how the expected PC incidence and mortality could have developed, in the absence of any screening. By comparing the observed and expected incidence and mortality in the screening group – with organized screening versus without any screening – NNI and NND for organized screening could be calculated. In the same way, by comparing the observed and expected incidence and mortality in the control group – with opportunistic screening versus without any screening – NNI and NND for opportunistic screening could be assessed. The study period was from Jan 1, 1995 to Dec 31, 2012. NNI and NND for organized screening were calculated for the last seven years of follow-up, but NNI and NND for opportunistic screening could only be calculated after 15 years, as this was the first time the observed mortality was lower than the expected.

Paper IV

This paper investigated the effect of age at screening and the number of screening visits (“screens”) on the risk of being diagnosed with PC. The screening arm of the Göteborg screening study contributed the study population (Figure 7). Men in the screening arm were defined as “attendees”

if they attended at least once and as “complete attendees” if they accepted all screening invitations without any interval interruption. Men who never attended were defined “non-attendees”. Complete attendees who were diagnosed with PC, who emigrated or who deceased between screening rounds (within 2 years of their last screen) were also included in the group of complete attendees. The maximum number of screens in each age cohort is shown in table 7.

Age cohort	Maximum number of screens
1930-31	3
1932-33	4
1934-35	5
1936-37	6
1938-39	7
1940-41	8
1942-43	9
1944	10

Table 7. Maximum number of screens for each age cohort

The study period was from Jan 1 1995 to Jun 30, 2014. Men born 1944 were excluded from analysis in the **paper IV** due to short follow-up. In addition, the screening algorithm in the 10th round, which only included men born 1944, was different from the previous rounds, with MRI used as a complementary screening test. Prevalent cases of PC (who were excluded from invitation, figure 7) were not evenly distributed in the age cohorts but were more frequent in the older age cohorts. We therefore added these cases to the each corresponding age cohort when calculating cumulative incidences of PC among complete attendees. Data regarding the number of PC deaths in the male Göteborg population born 1930-44 between 1980 and 1994 was obtained from the National Board of Health and Welfare.

3.3 Statistical considerations

In the four papers that constitute this thesis several different statistical methods have been applied. These methods are discussed here below, and a summary of the methods used in each paper is presented in table 8.

The Wilcoxon's rank-sum test/Mann-Whitney U test was used to test differences in medians between groups for variables that were not normally distributed, for example median age at diagnosis or death in **paper I**. **In paper I**, Cohen's kappa coefficient was calculated to measure agreement between COD certificates and the COD committee. The Kappa statistics is used to calculate agreement between two or more independent observers. As compared to a "simple" calculation of the proportion of agreement, the Kappa statistic takes into account that agreement can occur by chance. The calculation is based on the difference in the observed agreement compared to how much agreement could be expected by chance alone. Kappa can range from -1 to1, where 1 is perfect agreement, 0 is what would be expected by chance and a negative value would indicate systematic disagreement. A value between 0.81 and 0.99 is generally considered almost perfect agreement.[271] In **paper I**, confidence intervals (CI) for proportions (agreement, sensitivity and specificity) were calculated according to the efficient-score method with correction for continuity.[272, 273] The CI for the kappa statistics was calculated using an analytical method for dichotomous variables.[274]

Survival analyses

Different forms of survival analyses have been used in all four papers. Survival analysis is a branch in statistics that focuses on the time it takes for some event to occur, it other words, survival analyses typically deal with so called "time to event" data. These types of analyses enable study participants to be followed for various lengths of time and allow so called "censoring". Survival times are censored if the period of observation is cut before the event/endpoint could potentially occur. Reasons for censoring can be that individuals are lost to follow-up, emigrate or die from unrelated causes, since these individuals are no longer "at risk" for the outcome of interest if they are no longer alive, or we would never be able to accurately find out whether they experienced the event or not if they were lost to follow-up or emigrated and we lost track of them. These individuals can contribute with valuable information regarding time at risk even though they have not experienced the event of interest, since we know for certain that they were event-free during the time that we could follow them.[275]The different forms of survival analysis used in this thesis are explained in brief and discussed below.

The Kaplan Meier estimator of the survival function: The Kaplan Meier estimator or the product limit estimator is the probability of surviving or being event-free at a certain time considering time in many small intervals.[275] For each time interval the survival probability is calculated as the number of persons being event-free divided by the number of persons at risk. Those who are censored are no longer at risk. The cumulative probability of being event-free at a certain point in time is a chain of conditional probabilities where the probabilities of being event-free at all time-intervals preceding that point are multiplied.[275] There are some assumptions that should be met when using the Kaplan Meier method; those who are censored should have the same survival probability as those who continue to be followed and the survival probability should not be dependent on the time of recruitment (those who are recruited early have the same survival probability as those recruited later on). In addition, it must be possible to register the date of the event.[276] Survival analysis can also be used to estimate the probability of having experienced an event at a certain time. The cumulative probability or the cumulative incidence of an event can be estimated as 1 – the Kaplan Meier estimator. To statistically test whether there is a statistical difference between the groups in the probability of the event at a certain time point, the log rank test can be applied. It is a hypothesis test where the null hypothesis is that the survival distributions are equal at all follow-up times. It compares the observed number of events in each group to the number of expected events if the survival function were the same, taking in to account differences in the length of follow-up between the groups.[277]

-In paper I the cumulative incidence of death due to all causes, death due to PC and death due to non-PC causes in the screening and control group were plotted as 1 – the Kaplan Meier estimator. Time was calculated from date of diagnosis to date of death, emigration of last follow-up (Dec 31, 2008), whichever came first. The log rank test was used to test if there were significant differences the probability of death (from all causes, PC and other causes) between the screening and control group.

-In paper III median time to PC diagnosis in the screening and control group was calculated with the Kaplan Meier method. Time to diagnosis was calculated from study start (Jan 1, 1995) to date of diagnosed. Men who did not experience an event were censored at date of death, emigration or Dec 31, 2012, whichever came first. The log rank test was used to test the difference in time to diagnosis between the study arms.

Actuarial or life table analysis: This type of analysis is similar to the Kaplan Meier analysis but is used to describe data where the results are grouped into time intervals instead of the exact time that the events occur, as in the Kaplan Meier analysis. For the calculation, the number of persons at risk who enter the time interval, the number of events and the number censored during the interval are needed. The cumulative incidence can then be estimated as 1 – the actuarial survival estimate. Life tables can also be used to estimate the survival curve for a cohort of people from birth using age and sex specific mortality rates.[275]

-In paper III life-table analysis was performed to estimate the observed cumulative PC incidence and mortality in the screening and control group, calculated as 1 – the actuarial survival estimate. Follow-up time was calculated from study start (Jan 1, 1995) to date of an event (PC diagnosis or death). Men who did not experience an event were censored at date of death, emigration or last follow-up (Dec, 31 2012). Standard errors (SE) were calculated according to a method by Greenwood [278] and CI were calculated on the log cumulative hazard scale and then transformed back to the survival scale.

Cox proportional hazard regression analysis: This type of analysis can be used to explore the relationship of multiple predictors to a right-censored, time-to event outcome.[277] The hazard function is related to the survival function and the hazard can be interpreted as the instantaneous risk of having an event, assuming being event-free up to the time of interest.[275] The Cox proportional hazard model can test the independent effect of a number of explanatory variables, which can be continuous or categorical, on the hazard of the event. However, this analysis assumes that the relative hazard (the ratio of the hazards between the groups that are compared) is constant over time; the "proportional hazards assumption". This can be tested and addressed if the assumption is violated.[277]

Survival analysis in the presence of competing risks: A fundamental assumption in standard survival analysis, as mentioned above, is that censoring is not associated with an altered chance of having the event. Events that cause censoring and which are associated with an altered chance of experiencing the event are called competing risks.[279, 280] In the presence of competing risks the cumulative incidence calculated with the Kaplan Meier method (1 – Kaplan Meier estimator) will overestimates the cumulative incidence of the event; the magnitude of overestimation depends

on the incidence rate of the competing risk. For competing risk data, the log rank test is also inappropriate but corresponding hypothesis test for competing risk data exists. The Cox regression analysis also becomes invalid when competing risks are censored. There are competing risk regression models, that can distinguish between individuals who are still alive and those who have failed due to competing risks. The most common competing risk regression is the one by Fine and Gray.[281]

-In paper II the cumulative incidence of treatment and failure were calculated with both Kaplan Meier method and competing risk analysis in order to estimate the magnitude of overestimation with the Kaplan Meier method. In paper II, we also investigated the association between risk group and age at diagnosis with the hazard of failure after active surveillance. Due to the presence of competing risks, we performed both a Cox regression analysis and competing risk regression according to the method by Fine and Gray.[281] Follow-up time was calculated from the date of diagnosis to date of the event. Men who did not have an event were censored at the date of death (in the Kaplan Meier analysis and Cox regression), date of emigration of date of last clinical follow-up. For men with multiple events, i.e. PSA recurrence, PC metastases and PC death, time was calculated to the first occurring event.

-In paper IV, cumulative incidence of PC in the different age cohorts and at different ages was calculated as 1 – the Kaplan Meier estimator and with corresponding competing risk estimates. Time was calculated from study start to date of PC diagnosis. Men who were not diagnosed with PC were censored at date of death (in the Kaplan Meier analysis), date of emigration of last follow-up (Jun 30, 2014).

Ederer II method: In paper III, we sought to project PC incidence and mortality in the absence of opportunistic or organized screening. This was performed by calculating the expected PC incidence and mortality with something called the Ederer II method, based on historical data from years 1990 to 1994.[282] These years constitute the pre-PSA era when the usage of PSA as a screening tool for PC was close to non-existing. The Ederer II method is commonly used to assess expected survival when calculating relative survival or excess mortality. The expected survival can be thought of as being calculated for a cohort from the general population that is matched to the study cohort.[283] In paper III, the expected cumulative PC incidence and mortality were estimated as 1 – the Ederer II estimator. To match the age-distribution in the screening and control group when calculating the expected estimates we used 1-year age strata. The calculations of the

expected incidence and mortality were based on the entire male population in Göteborg. Alternative methods for calculating expected survival are Ederer I and the Hakulinen method. The three methods differ regarding how long each matched individual is considered to be at risk for the purpose of calculating expected survival.[283]

Number needed to screen (NNS) and number needed to treat (NNT): In paper III, we calculated NNS and NNT for organized and opportunistic screening. The statistics "number needed to screen" was first described by Rembold in 1998 and is a further development of the NNT by Cook and Sackett.[284, 285] NNS is calculated as 1 divided by the absolute risk reduction of the endpoint of interest between the study arms. In our study, NNS was calculated from an intention-to-treat analysis, that is, a comparison between all men in the two randomized arms, regardless of whether men in the screening group actually attended screening or not. Therefore, NNS referred to the number of men needed to be *invited*, and thus NNS was renamed number needed to invite (NNI). Number needed to treat (NNT) in a screening trial can be calculated as 1 divided by the absolute PC mortality reduction multiplied by the excess PC incidence. In the case of PC screening, a large proportion of those diagnosed with screen-detected cancer do not receive immediate treatment and NNT thus rather reflects the number of men needed to *diagnose*. NNT was therefore renamed number needed to diagnose (NND). Other authors have called this statistic number needed to manage (NNM) for the same reason. For studies with time to event outcomes, the calculation of NNS needs to take into account different length of follow-up and censoring, especially if censoring varies between the study groups. This can be done by calculating NNS from for example survival estimates from a Kaplan Meier analysis or the cumulative hazard function from the Nelson Aalen estimator.[286-288] In **paper III**, we calculated NNI and NND for organized and opportunistic screening, respectively. NNI for organized screening was calculated as 1 divided by the absolute difference in cumulative PC mortality between the observed PC mortality in the screening group compared to the expected pre-PSA era projected PC mortality rate. For opportunistic screening, NNI was calculated in the same way but by calculating the difference in the observed PC mortality for the control group compared to the expected pre-PSA era projected PC mortality rate. The absolute difference in PC incidence was calculated in the exact same fashion. NND was then calculated as NNI multiplied by the excess (increased) incidence between screening compared to pre-PSA era projected rates and control compared to pre-PSA era projected rates. In these calculations, cumulative PC incidence and mortality were calculated with the life-table/actuarial method and with the Ederer II method. Confidence intervals

for NNI was calculated with a method described by Altman et al.[288] Both NNI and NND are highly time dependent time dependent and citing only one estimate can be misleading, without also mentioning the follow-up time at which these metrics are calculated.[178]

In paper III, when calculating confidence intervals for the absolute and relative risk reduction the SE for the expected values was assumed to be zero as the expected incidence was based on the whole male population in Göteborg. The 95% CI for the absolute risk reduction was calculate as expected – observed cumulative incidence +/- 1.96 x $SE_{observed}$. The 95% CI for the relative risk reduction was calculated as 100 x (1 minus the quote between the limits of the 95 % CI for the observed and the expected cumulative incidences.

	Follow-up period	Study population	Applied statistical methods
Paper I	Jan 1, 1995 – Dec 31, 2008 (additional follow-up presented in this thesis up to Dec 31, 2012)	Men diagnosed with PC and who had deceased (screening and control arm)	Cohen's kappa score, Wilcoxon rank-sum test, Kaplan-Meier method (1-Kaplan Meier estimator), the logrank test
Paper II	Jan 1, 1995–Dec 31, 2010	Men with screen-detected PC managed with active surveillance	Kaplan Meier method (1 – Kaplan Meier estimator), Cox proportional hazards model, cumulative probabilities calculated with competing risks analysis, competing risk regression
Paper III	Jan 1, 1995-Dec 31, 2012	Men in the screening and control arm in the Göteborg screening trial, historical comparison based on the male population in Göteborg 1990-94	Kaplan Meier method, log rank test, Ederer II (1 – Ederer II estimator) actuarial/life table method (1 – actuarial survival estimate, NNI and NND
Paper IV	Jan 1, 1995 – Jun 30, 2014	Men in the screening arm in the Göteborg screening trial	Kaplan Meier method (1 – Kaplan Meier estimate), cumulative incidences calculated with competing risk method

Table 8. Summary of statistics, study population and study periods in paper I-IV

4 RESULTS

Paper I

This paper investigates the accuracy of Swedish COD certificates for men with PC to ensure that they can be used for determining the COD in the Göteborg screening study. During the follow-up of this study (Jan 1, 1995 to Dec 31, 2008), a total of 285 men diagnosed with PC deceased. Of these men, 278 had analyzable COD certificates and were included in the analysis in **paper I**.

Men in the screening arm were significantly younger at diagnosis compared to controls (64.9 versus 66.0 years, p=0.004) but there was no significant difference in the median age at death between the study groups (69.8 versus 70.3 years, p=0.89). As men in the screening group were diagnosed at a younger age, median follow-up from date of diagnosis was longer for the screening group than for the control group (6.7 versus 4.3 years, p<0.001). The vast majority of cases had a morphologically diagnosed PC (96%). Multiple malignancies (28%) and serious comorbidity were common (55%). Almost half (48%) of all deaths occurred in a hospital and 17 percent of the deaths resulted in an autopsy (medical or forensic).

The COD, as coded by the COD committee, is shown in table 9.

The certificates and the committee were in agreement regarding the COD in 267 of 278 cases, corresponding to an overall agreement of 96% (95% CI 92.8-97.9). Agreement, sensitivity, specificity, and kappa score are listed in table 10.

Definition of COD	Number of cases, follow up until Dec 31, 2008	Number of cases, follow up until Dec 31, 2012	In the final evaluation classified as
Definitely PC	119 (1)	197 (2)	PC
Probable PC	1	1	PC
Definitely intervention-related	1	1	PC
Possible PC	2	2	Other COD
Unlikely PC	1 (1)	4(1)	Other COD
PC as a contributory COD	4	19	Other COD
Definitely not PC	150 (5)	314(11)	Other COD
Pending	0	(1)	
Total	278 (7)	538 (15)	

Table 9. Coding of cause of death from the cause of death committee with follow-up until Dec 31, 2008 and Dec 31, 2012. (Within parenthesis, number of men who lacked an analyzable death certificate but where cause of death was possible to establish for the committee based on information from medical records).

	Agreement, follow-up to Dec 31 3008			Overall agreement, follow-up Dec 31 2012
	Overall	Screening group	Control group	
Agreement	96% (267/278)	97% (143/147)	95% (124/131)	94% (508/538)
Kappa score	92%	94%	89%	88%
Sensitivity	96%	98%	95%	92%
Specificity	96%	97%	94%	96%

Table 10. Measurements of agreement

There were 11 cases of disagreement, four cases in the screening group and seven cases in the control group. However, disagreement went in both directions with an incorrect PC as the COD in five cases and an incorrect other COD in six cases. The discordant cases were a diverse group and it was difficult to find any feature that distinguished them from the remaining cases.

To further validate the COD certificates, the cumulative overall-, PC-specific- and non-PC-specific mortality were estimated with the Kaplan Meier method. The screening group appeared to have a lower overall mortality compared to the control group but this difference was probably due to a lower PC mortality as the cumulative risk of dying from non-PC causes was similar in the two groups.

An updated review of the COD committee's work until Dec 31, 2012, which was not available at the time of publication of **paper I** has been added to tables 9, 10 and 11. With an extra four years of follow-up, the total number of men who had died with a PC diagnosis amounted to 553 men, of whom 538 had analyzable death certificates and were thus available for a comparative analysis. Median age at diagnosis in this population was 66.3 years (screening group 65.3 years, control group 67.4 years, $p<0.001$) and the median age at death was 72.9 years (screening group 73.1, control group 72.7, $p=0.98$). In total, there were 30 cases of disagreement up to Dec 31, 2012, corresponding to an overall agreement of 94%. There were 15 discordant cases in the screening arm and 15 cases among controls. There

were 17 cases with PC inaccurately assigned as the COD on the certificate where the committee had assigned other COD, and 13 cases where the COD was inaccurately coded as other COD on the certificate. The number of PC deaths in the screening and control arm as assigned by the COD certificate and the committee respectively, is shown in table 11.

	Follow-up until Dec 31, 2008		Follow-up until Dec 31, 2012	
	Certificate	Committee	Certificate	Committee
Screening group	45	43	83	78
Control group	77	78	120	121

Table 11. Number of prostate cancer deaths in each study group according to cause of death certificates and committee

The members of the committee were in total agreement regarding the COD in 84% (466/553). In 12% (65/553) there was minor disagreement between the members and in 4% (22/553) there was major disagreement.

Paper II

This paper investigated outcomes for men managed with AS in the Göteborg screening study.

In the Göteborg screening trial, almost half (45.7%, 442/968) of all men with a screen-detected PC were primarily managed with active surveillance. Three men were excluded; two men moved to another area in Sweden and another man refused active treatment and follow-up, resulting in a study population of 439 men. Median age at diagnosis was 65.4 years and median follow-up from diagnosis was 6.0 years (range 0.08-15.1 years). Of all screen-detected cases, very low-risk and low-risk constituted 60%, and in the active surveillance cohort these groups constituted 78%. Men with intermediate risk PC were a relatively large group and amounted to 21% of those on active surveillance.

During follow-up, a total of 162 men switched treatment strategy from active surveillance to radical prostatectomy (n=106), radiotherapy (n=32) and hormonal treatment (n=24). The 5- and 10-year cumulative risk of receiving

treatment were 38.5% and 54.6% with Kaplan Meier estimates (median time on active surveillance 8.2 years) and the corresponding figures for competing risk estimates were 37.5% and 51.4%, respectively. An increase in cancer involvement or Gleason grade was the most common reason for discontinuing active surveillance (48%), followed by an increase in PSA-level (28%), T stage progression (4.3%) and anxiety (2.5%). Ten men (6.2%) had unknown reasons for switching treatment strategy and 15 (9.3%) men deferred treatment on their own. During follow-up, only one man died due to PC and 59 men died from non-PC causes, corresponding to 10-year Kaplan Meier estimate for overall survival of 81%. Thirty-nine men were classified as failures; one man died from PC, another man developed metastasized PC, 23 men initiated hormonal treatment and 14 men experienced PSA recurrence after radical prostatectomy (n=10), radiotherapy (n=3) or radical prostatectomy plus salvage radiation (n=1). Of those who were classified as failures, 11 of 39 died from non-PC causes during the study-period. The 5- and 10-year cumulative risk of failure were 6.7% and 14% with Kaplan Meier estimates and the corresponding figures for competing risk estimates were 6.4% and 13%.

Risk group was significantly associated with failure in both univariate and multivariate Cox proportional hazard models as well as in the competing risk model. Compared with very low-risk PC, men with low-risk PC had a HR for failure of 2.1 and men with intermediate risk PC had almost four times the risk of failure (HR 3.7). Age was significantly associated with failure in the univariate analysis but this associating was no longer significant in the multivariate analysis or the multivariate competing risk model.

Paper III

In this paper we investigated the effectiveness of organized compared to opportunistic screening in reducing PC mortality, measured as NNI, as well as risk of overdiagnosis, measured as NND.

The total number of men diagnosed in the screening and control arm, and the method of diagnosis is listed in table 12. In the screening group, 87% were compliant with the recommendation of having a biopsy after a positive screen. Tumors detected by opportunistic screening were more advanced than those detected with organized screening (Table 12). Men in the screening group were younger at diagnosis than men in the control group (65.8 versus 67.8 years, $p<0.001$).

	Screening group		Control group	
Number of prostate cancer cases	1396		962	
	Number of screen-detected (organized screening) prostate cancer cases	Number of prostate cancer cases, not screen-detected	Number of screen-detected (opportunistic screening) prostate cancer cases	Number of prostate cancer cases, not screen-detected
	1022*	374	361	601**
Low-risk†	613 (60%)	84 (22%)	128 (35%)	125 (21%)
Intermediate risk‡	331 (32%)	138 (37%)	168 (47%)	192 (32%)
High risk§	63 (6.2%)	73 (19%)	42 (12%)	127 (21%)
Advanced¶	13 (1.3%)	54 (14%)	10 (2.8)	107 (18%)
Unknown	2 (0.2%)	25 (6.7%)	13 (3.6%)	50 (8.3%)
Number of prostate cancer deaths	79		122	

*Table 12. Prostate cancers diagnosed in the study groups *includes nine cases detected as a result of an erroneous invitation; **includes eight cases diagnosed at autopsy; †T1, not N1 or M1, Gleason score ≤6, and PSA < 10ng/mL; ‡T1-2, not N1 or M1, and Gleason score ≤7 and/or PSA < 20 ng/mL; §T1-4, not N1 or M1, and Gleason score ≥8, and/or PSA <100 ng/mL; ¶N1 and/or M1 and/or PSA ≥100 ng/mL*

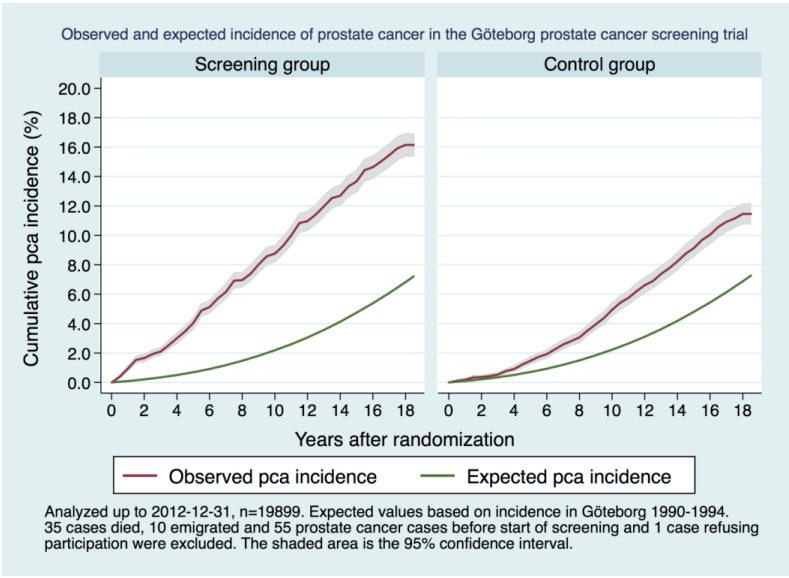

Figure 8. Observed and expected cumulative prostate cancer incidence in the screening and control group

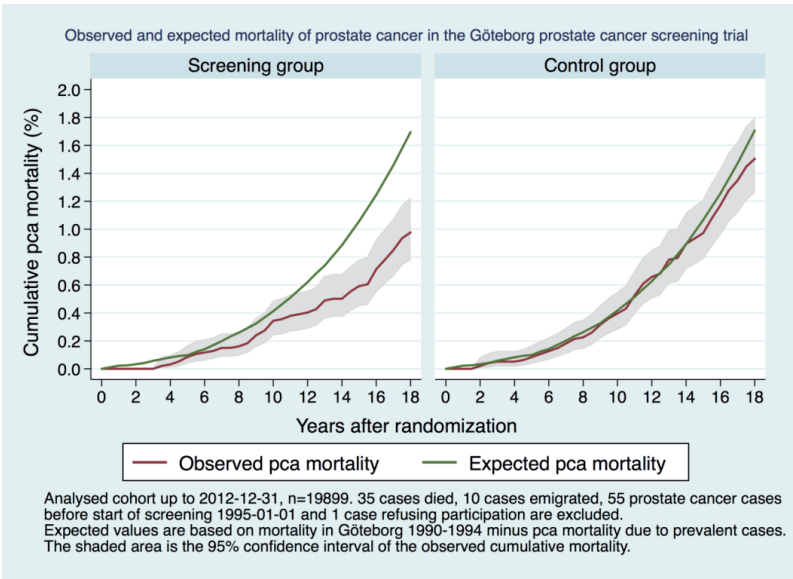

Figure 9. Observed and expected prostate cancer mortality in the screening and control group

The observed and expected PC incidence and mortality and the corresponding NNI and NND for the screening and control group after 18 years of follow-up are listed in table 13.

	Expected prostate cancer incidence	Observed prostate cancer incidence	Expected prostate cancer mortality	Observed prostate cancer mortality	
Screening group (organized screening)	6.8%	16%	1.7%	0.98%	NNI 139 NND 13
Control group (opportunistic screening)	6.9%	11%	1.7%	1.5%	NNI 493 NND 23

Table 13. Observed and expected prostate cancer incidence and mortality and number needed to invite and diagnose in the screening and control arm at 18 years of follow-up.

Organized screening, as performed in our screening arm, every second year with a PSA cut-off of 2.5-3 ng/mL for biopsy, caused a pronounced PC mortality reduction when compared to a situation without screening, that is the pre-PSA era expected PC mortality rate (absolute reduction 0.72%, 95% CI 0.50-0.94%, relative risk reduction of 42%, 95% CI 28-54%). In contrast, opportunistic screening, as performed in our control group including PSA-testing as part of clinical routine and as part of health check-ups, was not associated with a statistically significant PC mortality reduction compared to a situation without screening, that is the pre-PSA era expected PC mortality rate (absolute reduction 0.20%, 95% CI 0.06-0.47%, relative risk reduction of 12%, 95% CI -5-26%).

Paper IV

This paper aimed at investigating the effect of age and number of screens on the cumulative risk of being diagnosed with PC. With a follow-up until 30 June, 2014 there were 3485 (38%) men who were complete attendees, 3473 were attendees and 2107 were non-attendees. Of those with a positive screen, 87% complied with the biopsy recommendation. A total of 664 (19%) cases of PC (screen-detected and interval cancers) were detected among complete

attendees and 333 (9.6%) cancers among attendees. For the entire cohort, the median age at diagnosis was 65.1 years compared to 64.9 years for complete attendees and 65.8 years for attendees. The cumulative PC incidence increased steadily with age (Figure 10, Table 14). When men reached the upper age limit for invitation (average 69 years) there was no significant difference in PC incidence between age cohorts with the exception of the oldest cohort compared to the youngest (Table 14). The cumulative risk of PC reached a "steady state" after four screens and thereafter only increased with age and not with additional screens (Table 14). Similar results observed if a competing risk model was used. No obvious influence on PC incidence was observed when PSA threshold was lowered or the number of cores was increased (Figure 11).

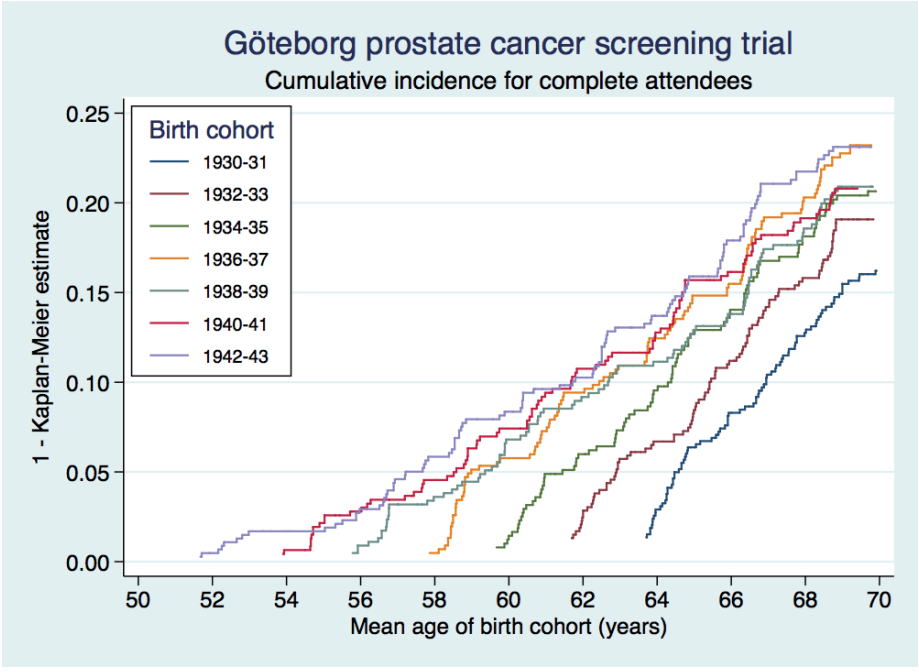

Figure 10. Cumulative incidences of prostate cancer for complete attendees divided by age cohorts (prevalent cases of prostate cancers are added to each cohort)

Birth cohort	Number of men in the cohort	Cumulative incidence at 60 years (number of screens)	Cumulative incidence at 65 years (number of screens)	Cumulative incidence at 70 years (number of screens)
30-31	1108		6.4 (1)	16.2 (3) (95% CI: 13.5-19.4)
32-33	1106		8.5 (2)	18.2 (4) (95% CI: 16.0-22.7)
34-35	1083	1.4 (1)	12.7 (3)	20.6 (5) (95% CI: 17.2-24.7)
36-37	1190	5.8 (1-2)	14.8 (4)	23.2 (6) (95% CI: 19.6-27.3)
38-39	1409	6.8 (2-3)	12.9 (5)	20.9 (7) (95% CI: 17.5-24.9)
40-41	1453	7.4 (3-4)	15.7 (6)	20.8 (8) (95% CI: 17.1-24.9)
42-43	1716	8.4 (4-5)	15.9 (7)	23.1 (9) (95% CI: 19.5-27.2)
Cumulative incidence for those with ≥4 screens		7.9% (95% CI: 6.2-10.1)	15.0% (95% CI: 13.5-16.7)	21.2% (95% CI: 19.7-2.8%)

Table 14. Cumulative risks for prostate cancers at different ages and number of screens among complete attendees, divided in age cohorts. Number of screens within parenthesis. Prevalent cases are added to each cohort.

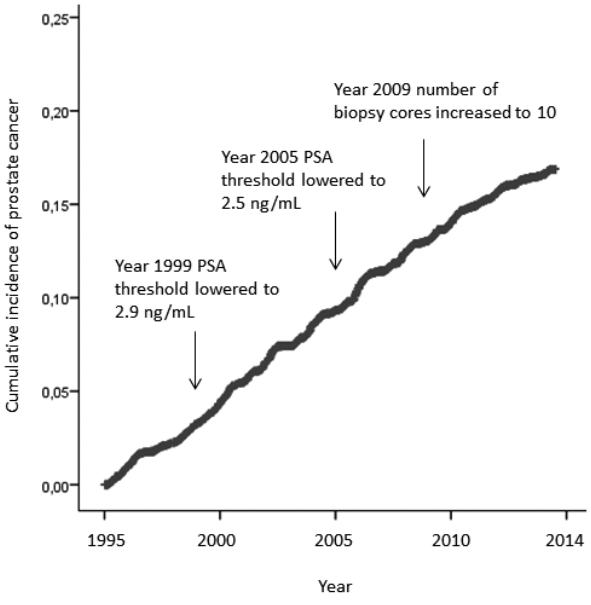

Figure 11. Cumulative incidence of prostate cancer in the screening group of the Göteborg screening trial.

5 DISCUSSION

5.1 Paper I – Cause of death determination

In cancer screening studies, cancer-specific mortality generally only constitutes a small proportion of the overall number of deaths. Therefore, trying to compare any differences in overall mortality would require extremely large sample sizes. However, using disease-specific mortality rests on the assumption that the COD determination is valid. It is therefore critical that the COD is assigned as accurately as possible and with minimal bias. To achieve this, many screening studies have used a COD committee reviewing medical records to determine the COD.[289-293] However, this method is expensive and time-consuming and would not be necessary if COD certificates were accurate. In the **paper I** of this thesis we therefore compared COD from Swedish COD certificates for men with PC in the Göteborg screening trial with the COD as assigned by an expert committee using a standardized algorithm. The results showed an excellent level of agreement at 96% during the first 13 years of follow-up. A high level of agreement was seen also after 17 years (94%). Similar estimates have been reported from other ERSPC centers using the same algorithm. For example the Dutch and the Finnish centers have reported an overall accuracy between official vital statistics or COD certificates and a COD committee of 90.6% and 97.7% respectively.[294, 295] One prior Swedish study by Fall et al. has investigated the accuracy of COD certificates for men with PC. In this study, overall agreement between COD from certificates and COD as determined by review of medical records was 86%. Agreement was higher for younger men and those with localized disease.[296] The lower level of agreement in the study by Fall could possibly be explained by an older population compared to the population in **paper I**.

In **paper I** the discordant cases were evenly distributed in the two study arms (15 in each group) but with the committee's decision as reference, more cases were inaccurately coded as PC deaths by the certificate (n=17) than the number of cases that were inaccurately coded as other COD (n=13). The difference in the number of PC deaths between the screening and control arms was larger when COD was determined by the committee then when based on certificates (Table 11). This difference increased between 13 and 17 years of follow-up, as evident by the decrease in sensitivity of the certificates (from 96% to 92%, Table 10). There are several potential explanations to these findings. There is obviously the possibility that this distribution has

occurred by chance. However, a few other studies have also investigated the effect of the mortality review on the main result of the trial. In the majority of these studies, the rate ratio of disease-specific mortality between trial arms has shown a larger benefit for screening when review data from a committee has been used compared to when data from death certificates has been used.[290, 292, 293] An alternative explanation, other than chance, is that the committee's COD assignment could be biased towards overestimating the effect of screening on disease-specific mortality relative to controls. Other screening studies have been criticized for having trial investigators as members of the COD committee and for insufficient blinding of medical records so that it has been possible to determine the study arm of the participant.[234, 290] In the present study, the committee consisted of three experience urologists who had no other involvement or commitment in the Göteborg screening trial. Even with efforts trying to blind the study arm for the committee, complete blinding is difficult to obtain. For example, a PC that is diagnosed at an early stage in an asymptomatic man is naturally more likely to belong to the screening arm then a man who is clinically diagnosed due to pain from bone metastases. The use of a standardized algorithm should have reduced the risk of a biased COD ascertainment compared to if the COD was subjectively ascertained through reviewing medical records. The algorithm was based on a number of clear questions and criteria that were needed to be answered and fulfilled.[268] Also, if the information from the medical records were not sufficient, the COD could be assigned "pending". This only occurred in one out of 553 cases.

A third possible explanation is a phenomenon referred to as "sticky diagnosis bias".[297] This bias is connected to overdiagnosis and causes disease-specific mortality between the screened group and the control group to be biased *against* screening (reducing the benefit). With screening, more men will be diagnosed with PC in the screening arm compared to the control arm. There is a risk that the COD will be coded as PC on the certificate even though these men died from something else ("sticky diagnosis"). In contrast, men dying from PC in the control arm could be falsely attributed to having died from other causes as these men are less likely to have been previously diagnosed with PC. As screening with PSA is associated with substantial overdiagnosis and leads to a large increase in the number of men receiving a PC diagnosis this bias could potentially be problematic and lead to difficulties in observing a reduction in PC mortality by screening compared to controls. The COD committee in the present study only reviewed COD for men with a PC diagnosis. There could be cases in the control group with unknown PC whose deaths were falsely attributed to other COD. However, these potential cases are likely to be very few. Among the discordant cases,

there were 10 cases in the screening group where the COD on the certificate was PC but where the committee assigned death from other COD. These cases could potentially be the result of sticky diagnosis bias. Misattribution of COD has in some studied been suggested to explain part of the PC mortality trends. [298, 299] However, there are also studies that have not found any change in the COD determinations since the introduction of PSA.[300] Another interesting finding, which is contrasting to the results from **paper I**, was reported from the Dutch center of the ERSPC. In their report on the death verification process, official statistics tended to over-report PC as the COD for men in the control arm.[294]

Whether the imbalance among the discordant in cases in our material is due to chance, inadequate blinding, sticky diagnosis or a combination of the three is impossible to find out. Anyhow, the most important finding is that the overall number of inaccurate COD certificates was small and their potential effect on the overall result of the Göteborg screening study would have been minimal.

Another bias that can occur in COD determination, that is also connected to overdiagnosis and overtreatment, is "slippery-linkage bias".[297] This bias would have an opposite effect; it would overestimate the positive effects of screening on PC mortality by not including intervention-related death in the definition of disease-specific mortality. As PC screening is associated with substantial overdiagnosis and overtreatment, it is important to try to identify this bias. In the Göteborg screening study, as well as in the entire ERSPC, intervention-related deaths are included in the definition of PC-mortality. In our material, there was only one case out of 553 deaths that was coded as intervention-related and whose death was "falsely" attributed to other COD on the certificate. This was a man in the control group who was diagnosed with a moderately differentiated PC with a PSA of 40 ng/mL at diagnosis. He was treated with external beam radiotherapy of a total 70Gy. Several years later he was diagnosed with bilateral hydronephrosis secondary to radiation fibrosis. He also had severe cardiovascular disease and deceased 12 years after the PC diagnosis in circumstances that could have been related to these treatment complications. One limitation with the COD determination process in our study is that the COD was only reviewed for men with PC. Potentially, there could be intervention-related deaths caused by the screening processes which could be missed. Consider, for instance, a man, who would have a fatal complication after a biopsy (e.g. sepsis) with benign result, his death would not be identified as intervention-related. However, fatal complications after prostate biopsy are very rare. Carlsson et al. investigated the possible excess mortality following a prostate biopsy in the Swedish, Finnish and Dutch

centers of the ERSPC. In their study, there was no significant difference in excess mortality for screening-positive men compared with screening-negative men. Fourteen men died within 120 days of the biopsy but none died as an obvious complication to the biopsy.[207]

A learning curve for the members in the COD committee was observed. With time, there was less usage of the more vague COD definitions such as "probable" and "possible PC", and a tendency to instead report COD as "definitely" or "definitely not" PC (Table 9). However, when looking at the cases where there was major disagreement between the committee members, cases of disagreement appeared to be evenly distributed over time and showed no tendency to decrease with time. In the absolute majority of cases 531/553(96%), the committee members were in agreement whether or not PC was the COD. The 22 cases (4%) for where there were major disagreement between the committee members tended to be cases where it was questionable whether PC had contributed to death or not. There was only one case where the committee members had reached completely opposite conclusions (2 members assigned death to definitely not PC 2 and 1 member to definitely PC).

Paper I has several strengths such as the use of a standardized algorithm and an independent COD committee without any other connection with the study. The Swedish personal identification number enabled linkage to other registers and facilitated the collection of medical records from outside sources such as primary care facilities. In the majority of cases, the COD committee therefore had access to complete medical records from the time of PC diagnosis. Even medical records from abroad were collected. All the collected material was also reviewed by first author to assess potential risk factors for an inaccurate COD certificate. There was a relatively high autopsy frequency (17.3%) which should improve COD determination as autopsy is the golden standard, and the most thorough method for determining the COD. The autopsy frequency is the proportion of deaths where the COD is determined with a clinical or forensic autopsy. In Sweden, the autopsy frequency has decrease from approximately 50% in the 1970s to 11% in 2012.[265]

Methodological considerations

There are some issues in **paper I** that deserves discussion; the compositions of the COD committee, the complete access to medical records, the recording of comorbidity status, the decision to only evaluate COD for men with PC and the age of the study population. First, it can be discussed whether the composition of the COD committee with three urologists could have affected

the results. Hypothetically, urologists could be biased to be pro- or against PC screening. However, there is no consensus regarding PSA screening in the Swedish urological community. Moreover, it is unclear whether a different composition of the committee would have changed the results. For example, in the Finnish section of the ERSPC the committee consisted of a specialist in forensic medicine, one urologist and one specialist in internal medicine. Yet they reached a similar level of agreement as in our study, between certificates and committee (97.7%). However, in 53% of cases, the members of the Finnish committee disagreed on the COD and consensus had to be reached through discussion. The corresponding figure in our material was 16%, indicating that COD coding in our committee was much more homogenous, which could be a result of the composition of the committee. Secondly, as previously discussed, it was difficult to accomplish complete blinding with respect to study arm, since the committee had access to complete medical records. If the COD committee was aware of trial arm allocation, they could, in theory, be more likely to assign PC as the final COD for a man in the control arm than for a man in the screening arm – or alternatively that the greater overdiagnosis and labeling of PC of men in the screening arm would lead to increased labeling of PC as COD for these men compared to the control arm. It has been suggested that the difficulties in blinding, could be improved if only records from a man's last life-year are reviewed. Whether this strategy would be superior or not is unknown but it could also be that such a strategy would miss identifying certain PC deaths such as late complications from treatment performed many years before. Third, as the review process of medical records is very time-consuming, only COD for men with PC was reviewed. This could have led an underestimation of the rate of "sticking diagnosis bias" and "slippery-linkage bias" in the COD certificates, as previously discussed. Fourth, comorbidities were not evaluated according to a validated index or score such as the Charlson comorbidity index [53] or the American Society of Anesthesiologists score (ASA score) but was subjectively assessed. However, this should not have affected the conclusions drawn from **paper I** but makes it difficult to compare the degree of comorbidities in this study population with estimates of comorbidity in other studies. Fifth, even with an additional four years of follow-up the study participants were relatively young when they died (median age 72.9 years) compared to the median age at death from PC in Sweden which in approximately 80 years. Because higher age, multiple malignancies and comorbidities are known risk factors for difficulties in COD determination, [296, 299] it can be hypothesized that the accuracy of COD certificates will decrease with time as the median age at death in the population increases.

In summary, the results from paper I and the additional analysis with added follow-up reported in the present thesis, show that COD certificate for Swedish men with PC are accurate, suggesting that they can be used for endpoint evaluation in PC screening studies with men in the same age range. However, in order to identify intervention-related death and for the group of patients lacking a death certificate (3%), having a COD committee is valuable.

5.2 Paper II – Reducing overtreatment

This paper investigated active surveillance as a management strategy to reduce overtreatment of screen-detected PC. The results showed that a large proportion of cancers that are detected by screening have very low-risk or low-risk features (60%) and could possibly be suitable for active surveillance. Almost half of all men with screen-detected PC in the Göteborg screening trial were primarily managed with active surveillance. For men with very low-risk and low-risk PC active surveillance appeared safe, at least in the medium-term, as no man died from PC or developed metastasized disease. For those with intermediate risk PC, active surveillance, came at a higher risk. These men had a four-fold higher risk of failure compared to those with very low-risk cancer. However, active surveillance could still be an option for some carefully selected men with intermediate risk (due to age and/or comorbidities) as long as they are well-informed that delaying treatment might risk their chance of cure. In this study of men with screen-detected PC, the risk of dying from unrelated causes was much higher than the risk of dying from PC. Only one man died from PC compared to 59 men who died from other causes.

The Swedish National Board of Health and Welfare (*sv*. Socialstyrelsen) recently published its updated 2014 National Guidelines for the Treatment Prostate Cancer. In these guidelines, the Board recommends that active surveillance should be the primary treatment option for men with very low-risk and low-risk PC. The Board set the aim that at least 95% of all men with very low-risk PC and a life expectancy of at least 10 years should be managed primarily with active surveillance.[301] The latest report from the NPCR showed that the usage of active surveillance in different counties (*sv*. landsting) in Sweden varied from 31% to 100% of all patients with very low-risk PC.[58] The same pattern was seen for men with low-risk PC. However, as comparing to international PC treatment trends, Sweden has a high usage of active surveillance and was early in adopting this management strategy. For example, in the US, the tradition has been to radically treat also men with

very low-risk and low-risk PC, and less than 10% of men with low-risk PC has been primarily managed with active surveillance.[27] A national survey among US radiation oncologists and urologists investigated the physicians' perception of active surveillance and showed that the majority believed that active surveillance was underused, furthermore, they felt comfortable recommending it. Despite this, only 22.1% recommended active surveillance for a case of a 60 year old man with low-risk PC. The physicians felt that their patients were not interested in active surveillance and the majority (56%) of radiation oncologists recommended radiotherapy (external beam radiation or brachytherapy) whereas most (61%) urologists recommended radical prostatectomy. In addition, they were concerned about overtreatment with radical prostatectomy and radiotherapy but radiation oncologists and urologists believed that primary treatment provided by the other specialty was overused.[302] From the CaPSURE register, which covers 40 urology practices in the US, Cooperberg et al. reported that only 6.8% of men with localized PC (diagnosed 1990-2008) were managed with watchful waiting or active surveillance and less than 9% of those with CAPRA score 0-2 (low-risk PC).[27] However, there seems to be a growing acceptance for active surveillance also in the US. From a register in Michigan, it was recently reported that 49% of patients with low-risk PC, according to the D'Amico criteria, diagnosed 2012-2013, were initially managed with active surveillance. However, the use of active surveillance varied greatly across practices in the register, ranging from 27% to 80%.[303]

There are several challenges with active surveillance, of which the first is to choose the right candidates. Choosing inclusion criteria for active surveillance is a balance between maximizing the number of men who can be eligible for active surveillance and thereby delaying or avoiding curative treatment but at the same time minimizing the number of missed potentially lethal tumors. The inclusion criteria of six large active surveillance-cohorts are listed in table 15. In a head-to-head comparison of five contemporary active surveillance criteria (Royal Marsden and University of Toronto not included) the Prostate Cancer Research International Active Surveillance (PRIAS) and the University of Miami criteria showed superior performance when it came to identifying insignificant PC and organ-confined GS 6 cancer in 391 radical prostatectomy patients. As expected, the most conservative criteria, the Johns Hopkins/Epstein criteria had the highest specificity but at the cost of a low sensitivity which resulted in a discriminative ability that what not superior to the other criteria. The PRIAS and the University of Miami criteria selected twice as many patients as the Johns Hopkins criteria but were still able to selected patients with the same pathological characteristics as the Johns Hopkins criteria.[304] In our study (**paper II**)

there were no pre-defined inclusion criteria for active surveillance but treatment strategy was chosen at the discretion of the treating physician and the patient. This resulted in a population not solely consisting of very low-risk and low-risk PC but also intermediate and high risk PC. This enabled us to study the effects of deferring treatment also for these groups. The results showed that there was a strong association between risk group and the risk of failure after active surveillance. The risk of failure was twice as high for men with low-risk tumors compared to men with very low-risk. Questions have been raised whether inclusion criteria for active surveillance should be expanded to also include men with intermediate risk.[305] So far, there is little data to support this. In our study, men with intermediate risk PC had a four-fold higher risk of failure compared to very low-risk patients. The one man who died from PC and the one patient who was diagnosed with metastasized PC in our cohort both belonged to the intermediate risk group. In the Toronto active surveillance cohort, 85 patients (19% of the entire cohort) had intermediate risk PC, and 58% (49/85) of those remained untreated during the study period (median 6.8 years). One man with intermediate risk PC developed PC metastases and died from PC.[306] The UCSF cohort has also reported outcomes for 90 men with CAPRA score 3-5 (intermediate risk). These men were not significantly different from those with CAPRA score 0-2 (low-risk) with regards to the proportion of men who were upgraded at repeat biopsy, the proportion who received curative treatment or progression-free survival. However, men with intermediate risk had higher rates of adverse pathological outcomes (positive surgical margins, pT3 and upgrading) after radical prostatectomy.[307] From the Finnish and Dutch centers of the ERSPC medium term outcomes for those who had deferred treatment were favorable for both low- and intermediate risk PC although treatment-free, metastases-free and disease-specific survival were somewhat lower for intermediate risk PC.[308] One argument for expanding the inclusion criteria could be that many men that are diagnosed with Gleason score 3+4 today would have been classified as Gleason 3+3 before the ISUP modification of the Gleason score in 2005. These men probably have a better prognosis than those who were classified as Gleason score 3+4 a decade ago.[37]

Institution	Inclusion criteria	Follow-up	Triggers for intervention
UCSF[307, 309-311]	Gleason score ≤3+3, PSA <10 ng/ml, ≤T2a, ≤33% biopsies +, ≤50% any core+	DRE, PSA 3 months, TRUS 6-12 months, repeat biopsy every 12-24 months (after 2003)	Increased Gleason score at repeat biopsy and/or PSAV increase >0.75 ng/mL/year, patient's desire
Royal Marsden Hospital[312, 313]	Gleason score ≤3+3 (≤3+4 if age >65 years), PSA < 15 ng/mL, T1-2, ≤ 50% biopsies+	DRE and PSA 3-6 months, repeat biopsy after 18-24 months and every 2 year	Cancer progression (Gleason score≤4+3 or >50% biopsies +) and/or PSAV >1 ng/mL/year
University of Miami[314]	Gleason score ≤3+3, PSA ≤10 ng/mL, T1-2, ≤2 biopsies+, ≤20% cores+	DRE and PSA 3-6 months, annual repeat biopsy	Cancer progression (any Gleason >3 or increase in number or volume of cores), patient's desire
Johns Hopkins[315]	Gleason score ≤3+3, PSA D <0.15 ng/mL, T1c, ≤2 biopsies+, ≤ 50% cores+	DRE and PSA every 6 months, annual repeat biopsy	Cancer progression (no longer meeting biopsy inclusion criteria)
MSKCC[316]	Gleason score ≤3+3, PSA <10 ng/mL, T1-2a, ≤3 biopsies+, ≤ 50% cores+	DRE and PSA (free and total) every 6 months, repeat biopsy after 12-18 months thereafter every 2-3 years	Cancer progression (no longer meeting inclusion criteria)
PRIAS[317]	Gleason score ≤3+3, PSA ≤10 ng/mL, PSAD <0.2 ng/ml/cc, T1-2, ≤2 biopsies+	PSA every 3-6 months, repeat biopsy year 1, 4 and 7 (number of cores volume dependent)	Cancer progression (Gleason 4 or 5, >2 biopsies+) and/or PSA-DT <3 years
University of Toronto[306, 318]	Gleason score ≤3+3, PSA ≤10 ng/mL (up to 1999 patients >70 years with Gleason 3+4 and PSA <15 ng/mL)	PSA every 3-6 months, repeat biopsy within a year and then every 3-4 years	Histological upgrading and/or PSA-DT <3 years and/or clinical progression

Table 15. Active surveillance cohorts and their criteria for inclusion, follow up and triggers for intervention. UCSF=University of California San Francisco. Biopsy += cores with cancer, cores+= proportion of core with cancer, PSAV=PSA velocity, PSA-DT= PSA doubling time

The second challenge with AS is to choose the "right" strategy for follow-up and "right" triggers for intervention. The aim should be to identify men with progressive disease without missing the window of curability. Similar to inclusion criteria, this is a balance between avoiding overtreatment of those who do not need treatment and identifying those who do. Most active surveillance protocols include an early confirmatory biopsy to identify potential misclassification at the first biopsy and repeat biopsies during follow up to detect disease progression (Table 15). The optimal number of biopsy cores, the location of these and the interval between repeat biopsies are currently unknown. The aim is trying to achieve a high degree of certainty that no Gleason 4 or 5 pattern has been missed. Several studies have shown that standard prostate biopsies often miss tumors in the anterior part of the prostate.[138, 319] The Swedish National Guidelines for PC Care therefore recommend that the anterior parts of the prostate are covered in the diagnostic biopsies before starting active surveillance [64]. Others advocate saturation biopsy (20-24 cores) or transperineal mapping biopsy (biopsies covering the entire gland at 5 mm interval) to rule out high-risk tumors in patients who are candidates for active surveillance. [138, 139]

Triggers for interventions differ slightly between cohorts. Some cohorts include PSA kinetics as a trigger for intervention (Table 15). The value of PSA kinetics, i.e. PSA velocity or PSA doubling time (PSADT), to monitor disease progression for low-risk PC is uncertain, and there is increasing evidence that it does not add predictive information regarding disease progression for men on active surveillance.[130, 320, 321] Reclassification to higher risk disease at repeat biopsy is common (17-55%), but with most series reporting estimates ranging from 20-30%.[135, 319-327] From the PRIAS study, Bul et al. reported that a repeat biopsy within a year (10 cores) resulted in unfavorable biopsy outcomes (increased Gleason score and/or increase in number of biopsies with cancer) in 21.5%.[328] Similar results have been reported from the MSKCC cohort where an immediate repeat biopsy within 3 months resulted in 27% of patients being upgraded or upstaged.[134] Bearing in mind the long natural course of PC and a considerable lead time for many screen-detected PCs, the majority of tumors being upgraded or upstaged at repeat biopsy likely represent misclassification rather than true disease progression. A positive confirmatory biopsy, PSAD and cancer involvement (number of cores and % invasion any core) have been reported to be associated with unfavorable histological finding at later repeat biopsy.[326-328]

In paper II, sextant biopsy was the standard up to 2008, thereafter a 10-core biopsy was used. There was no strict protocol defining re-biopsy interval.

These were typically performed if the first biopsy showed <2mm cancer, if there were signs of disease progression and every second and third year for men with stable disease. The most common reason for initiating treatment was an increase in Gleason score or cancer involvement followed by an increase in PSA. During follow-up approximately a third of the cohort (37%) discontinued active surveillance and received treatment (radical prostatectomy, radiotherapy, hormonal treatment). It is interesting to see that, despite that our study (**paper II**) did not have any strict inclusion or follow-up criteria this figure is very similar to the proportion of men who has discontinued active surveillance in other cohorts (Table 16). During the first two years of follow-up, there was a high incidence of discontinuing active surveillance. This probably indicated undersampling at the first biopsy rather than true disease progression.

Active surveillance holds many promises for the future and will most certainly be an important strategy to reduce to harms of overdiagnosis and overtreatment, at least until we have a screening method that can selectively identify only clinically relevant tumors, if one such strategy will ever be identified. Nevertheless, there are some questions remained to be answered regarding active surveillance. Inclusion criteria, follow-up strategies and triggers for intervention need to be validated and supported by evidence from long-term follow-up of ongoing large studies, but the most important question is probably: What "price" will men have to pay in order to delay or avoid the potential side-effects of curative treatment? There are several different aspects that have to be considered.

Firstly, what are the risks with repeat biopsies including a high number of cores? Complications after prostate biopsies include hematospermia, hematuria, rectal bleeding and infectious complications.[205] The rates of infectious complications following transrectal prostate biopsies have increased over time, mainly due to a higher prevalence of antimicrobial resistance.[202-205] A Canadian study reported that the risk of hospital admission within 30 days of a prostate biopsy increased 4-fold between 1996 and 2005.[202] The risks of repeat biopsies for men on active surveillance are under-studied. One study found that the number of previous biopsies was associated with an increased risk of infectious complications at subsequent biopsies (OR 1.33, 95% CI 1.01-1.74) whereas another study could not corroborate these findings.[329, 330] It has also been hypothesized that repeat biopsy can have a negative impact on potency but it has been difficult to separate out the negative effects of multiple biopsies from that of the normal ageing process on erectile function, as well as pshychological effects on potency of living with untreated PC.[331, 332]

Institution	Median follow-up years	N (median age, years, at diagnosis)	Proportion active treatment	Overall survival/mortality	Cancer specific mortality	Progression free survival	Treatment free survival
UCSF[307, 309-311]	5.0	810	43%	98% (5yr)	-	40% (5yr)	60% (5yr)
Royal Marsden Hospital[312, 313]	5.7	471 (66)	31%	96% (5yr)	2 PC deaths	NR	70% (5yr)
University of Miami[314]	3.7	230 (64)	14%	NR	NR	NR	86% (5yr)
Johns Hopkins[315]	2.7	769 (including 22% not meeting inclusion criteria) (66)	33%	14 deaths	-	≈65% (5yr)**	59% (5yr) 41% (10yr)
MSKCC[316]	1.8#	238 (64)	36%	NR	NR	60% (5yr)	NR
PRIAS[317]	1.6	2494 (66)	21%	87% (4yr)	-	NR	77% (2yr) 67.7 (4yr)
University of Toronto[306, 318]	6.8	450 (70)	30%	68% (10yr)	97.2% (10yr)	NR	72% (5yr) 62% (10yr)

*Table 16. Results from different cohorts. *median age at diagnosis, **read from figure, #median follow-up for patients without progression*

Secondly, what are the psychological consequences of living with an untreated cancer and how does active surveillance affect QoL? Many patients perceive a cancer diagnosis as a death sentence. In such a scenario, it is understandable that being recommended an expectant management of a potentially lethal disease may seem unreasonable. Physicians are faced with a difficult task trying to explain the rationale behind active surveillance. The physician's recommendation play a critical role is patient's acceptance of active surveillance.[333] Estimates of the proportion of men who are eligible for active surveillance and who actually choose this strategy have varied between 10-57%.[27, 334] In **paper II**, 59% of men with screen-detected very low-risk or low-risk PC were managed with active surveillance. Men managed on active surveillance is a highly self-selected group, which must be kept in mind when analyzing the reported levels of anxiety and QoL for men on active surveillance. In addition, the evidence available today regarding psychological effects of active surveillance comes from short-term results from non-randomized studies. However, these studies report low and favorable levels of anxiety and distress and maintained or even improved, levels of quality of life for men on active surveillance.[335, 336] The proportion of men who discontinue active surveillance without signs of disease progression has in most series been reported to be less than 10% [60, 306, 313, 315, 317] – but some series have reported levels of discontinuation of up to 20%.[337, 338] In **paper II** only four men (2.5%) reported that they came of active surveillance due to anxiety. Factors predicting adverse psychological effects include recent diagnosis, lack of a partner, previous mental health problems, consultations with fewer physicians prior to the treatment decision, and fewer cores at the diagnostic biopsy.[339] Strategies to support men and address their anxiety to improve adherence in active surveillance programs could be to include partners and loved ones in the process, working with the word "cancer" and offering patient support groups and contact nurses.[340] However, the evidence for these strategies is limited.[333]

Thirdly, does delaying treatment affect the possibility of nerve-sparing surgery and do men who choose deferred treatment have "more to lose" with regards to HRQoL when they are eventually treated? Deferring treatment until signs of disease progression could hypothetically lead to a more extensive resection at surgery with a concomitant smaller chance of a successful nerve-sparing procedure. This is, so far, an under-studied area partly because of methodological difficulties. One small, retrospective study compared HRQoL and sexual function scores for men treated with delayed radical prostatectomy after a period of AS with a group of men who underwent immediate radical prostatectomy. Men on AS who received

delayed treatment had more favorable HRQoL and sexual function scores preoperatively than men undergoing immediate surgery, however, this difference between the two groups could not be observed postoperatively suggesting that deferred radical prostatectomy had a larger unfavorable effect. It is interesting though, that in this small study there was no significant difference in the extent of nerve preservation, measured as fascia preservation score, between the two groups.[341] In contrast, another small, retrospective study reported significantly lower rates of bilateral nerve sparing surgery for men who deferred radical prostatectomy compared to those treated immediately.[342]

And finally, the most important question: does delaying treatment reduce the chance of cure? Cancer-specific mortality is the ultimate end-point for active surveillance studies but the slow-growing nature of low-risk PC with addition of several years of lead time means that it will take many years before this endpoint can be evaluated. Therefore, this remains an unanswered question as of today, and while awaiting long-term data, and we will have to rely on current evidence relating to intermediate outcomes such as disease progression, adverse pathology at radical prostatectomy, PSA recurrence after treatment and use of hormonal treatment. As previously mentioned, upgrading at repeat biopsy is common (20-30%). Regarding pathology results after active surveillance there are only a few, small, prospective series of radical prostatectomy following active surveillance. From the PRIAS study, pathological findings of 167 men who underwent radical prostatectomy after a median of 1.3 years on active surveillance have been reported. The vast majority had organ-confined disease (80.8%) Unfavorable radical prostatectomy results, defined as pT3-4 and/or Gleason score ≥4+3 was found in 29%.[343] Similar results have been reported in other series.[313, 344, 345]

Another strategy has been to retrospectively compare the result of immediate versus delayed radical prostatectomy in men with low-risk PC. Most series have found no significant difference in pathological outcome at radical prostatectomy between immediate versus delayed radical prostatecomty. [310, 346, 347]

A third strategy has been to study the outcome of radical prostatectomy in men who would have been eligible for active surveillance. For example, from the Swedish NPCR pathology outcomes of 4500 men with Gleason score 6 and T1c-T2 disease were reviewed to identify variables associated with adverse pathological outcomes. Approximately 50% had adverse pathological outcomes (pT3 or Gleason score ≥7). Of those patients who fulfilled the

inclusion criteria of six common active surveillance protocols adverse pathological outcome ranged from 33% to 46%. As could be expected, the most inclusive criteria (University of Toronto) resulted in the highest rates of upgrading (46%), whereas the lowest rates (33%) were found with the most stringent criteria (Johns Hopkins). In a multivariable model, age, total PSA, PSAD >0.15 ng/mL/cc, extent of cancer >4 mm in biopsies and palpable disease were predictors of adverse pathological outcome after radical prostatectomy.[348] A review by van den Bergh et al. summarized the evidence for delaying curative treatment for PC and concluded that treatment delay for months or even years did not seem to affect outcome for men with low-risk PC but that the quality of the evidence was weak. [349] However, there are several different aspects that should be considered when discussion these results. In the reported studies the maximum time of delaying treatment has been short (≈2years) and treatment might be deferred for many years for men on active surveillance. We do not yet know which future consequences this will have. Also, it should be kept in mind that adverse pathological outcome after surgery does not have to imply a poor prognosis in the long run. In addition, many of these studies are retrospective, most have different criteria for inclusion, follow-up and triggers for intervention and none is randomized. There is also a high risk of selection bias at multiple levels; those who accept active surveillance are probably not a representative sample of the general population and those on active surveillance who are treated with radical prostatectomy have had a reason to discontinue active surveillance (disease progression). To compare their outcome to men who are primarily treated with radical prostatectomy might be misleading. The intention with active surveillance is to identify those with disease progression within the window of curability. It is difficult to know whether worse pathological outcome for these men, in comparison to those primarily treated, is a failure or if the program has actually done what it intended, i.e. ensuring that those with aggressive disease are not left untreated. It might also be that among those who discontinue active surveillance, men who are treated with radical prostatectomy are likely those with the most favorable disease characteristics, whereas those with signs of more aggressive disease are more likely to be treated with radiotherapy or hormonal treatment.

Only a few series have reported PC deaths among men managed with active surveillance. In **paper II**, one man died from PC and another man developed PC metastases during a median follow-up of 6 years. The Toronto cohort, which has the longest follow-up, has reported, in a congress abstract at the EAU, that this cohort now has 14 PC deaths and another 16 patients who are alive but with PC metastases, among 840 patients managed with active surveillance with a median follow-up of 8.1 years. The actuarial 10-, 15-, and

20-year cancer-specific survival was 97.9%, 93.5%, and 86.3%. Of the patients who were treated, 21% had biochemical failure.[350] Medium-term outcome has also been reported from the Royal Marsden active surveillance cohort which have had two PC death among 471 active surveillance patient with a median follow-up of 5.7 years.[313] It is impossible to know if any of these deaths would have been prevented with immediate treatment. The evidence for immediate surgery for men with low-risk PC is uncertain and surgery for men with low-risk PC is difficult to justify.[68]

Methodological considerations

The main limitation in **paper II** was the lack of a protocol detailing the criteria for inclusion, follow-up and triggers for intervention. This hampers the possibility to compare the results from **paper II** with the results from protocol-based active surveillance studies. The lack of inclusion criteria resulted in a heterogeneous study population where some men where followed with watchful waiting rather than active surveillance. This is evident from the fact that the population also consisted of men with intermediate and high risk PC who traditionally are not regarded as candidates for active surveillance. The fact that 24 men started hormonal treatment during the study period is a further indication that the study population was no homogenous active surveillance cohort. However, from a clinical perspective, there is no clear-cut boundary between active surveillance and watchful waiting; as men grow older and gain comorbidities, active surveillance gradually transcends into watchful waiting.

The use of a sextant biopsy deserved mentioning. When the study started, this was the standard method, but more recent studies have shown that a sextant biopsy results in many false negatives and that a 10-12 core biopsy detects 30% more cancer.[140] The risk of undersampling is an important issue as it may lead to delayed of the correct risk group classification. With these limitations, it could be questioned whether the results from **paper II** are applicable to a group of carefully selected men with very-low or low-risk PC followed with a protocol-based active surveillance strategy. The results should instead be viewed as the outcomes of active surveillance for men in the general population and gives a picture on how active surveillance can be used in everyday practice. The results indicate that even with a population that is not heavily selected and followed according to carefully, pre-specified criteria the medium-term outcomes after deferred treatment are promising. The lack of a pre-defined follow-up protocol was probably not very influential on the results as the study population was mainly followed at the same clinic by a small group of urologists with the same treatment policies. This is strengthened by the fact that the proportion of men who were

managed with active surveillance was relatively constant during the study period, indicating that similar inclusion criteria were used. It is also interesting to observe that the results from **paper II** regarding the proportion of men discontinuing active surveillance and the reason for discontinuing are in line with results reported from cohorts with stricter criteria for inclusion and follow-up (Table 15 and 16). Despite having a follow-up that was longer than many other active surveillance cohorts, a median follow-up of six years is far too short to conclude that active surveillance is safe in the long term. The long-term findings from the Johansson cohort showed that localized PC can become lethal after having been stable for many years.[351]

5.3 Paper III and IV – Drivers of overdiagnosis

The results of **paper III** and **paper IV** have contributed with knowledge concerning factors that drive overdiagnosis and how screening can possibly be organized in order to reduce overdiagnosis. The results of **paper III** showed that opportunistic screening was far less effective in reducing PC mortality than organized screening and was associated with even more overdiagnosis when measured as NND. Organized screening resulted in a substantial mortality reduction (relative reduction of 42%) but there was no significant difference in PC mortality from opportunistic screening as carried out in Sweden at this time compared to pre-PSA era rates. In **paper IV**, the risk of being diagnosed with PC was heavily affected by age at screening but the number of screens had a minor effect. Furthermore, we could not identify any obvious association between changes in the PSA threshold during the course of the Göteborg trial or increase in the number of biopsy cores over time with the risk of PC diagnosis.

Drivers of overdiagnosis can be found both in the screened population and in the screening algorithm. As there is no easy way to directly measure overdiagnosis it is also difficult to measure the effects of various risk factors. The evidence is therefore, to a large extent, based on simulations from modeling studies.

Populations factors

Ethnicity, age and comorbidities affect the risk of PC and the risk of overdiagnosis. A recent review of international autopsy studies among men without a clinical diagnosis of PC confirmed previous findings that the prevalence of histological cancer increases with age, ranging from 2% from white men in their 20's to 69% of men above 90 years. African Americans

had the highest prevalence of PC and Asian men had the lowest prevalence when compared to white men. The size of the pool of prevalent cases indicates the upper bound of overdiagnosis, that is, how much additional cancer can potentially be detected.[20] In addition to having a high PC incidence, African American men are younger at diagnosis, are more likely to present with higher grade and stage and have an higher risk of dying from PC, compared to Caucasians.[352] Based on these findings, one would hypothesized that the risk of overdiagnosis should be lower among African American men. In contrast, several studies have reported that the overdiagnosis rate is, in fact, higher among blacks than whites.[163, 164]

Another population factor that is an important driver for overdiagnosis is age. Several studies, mainly modeling studies, suggest that age is a very important, perhaps the most important, driver of overdiagnosis.[19, 165, 172, 173, 198] Gulati et al. used a microsimulation model based on data from an American population, and modeled individual men's life histories in a virtual population. Subsequently, the data was fitted to a logistic regression model to construct a nomogram that could predict the risk of overdiagnosis based on age, Gleason score and PSA. Age at diagnosis was the most important predictor of overdiagnosis; for each additional year of age the odds of overdiagnosis increased by 12.9%. Gleason score ≥7 was associated with a decrease in the odds (19.5% compared to Gleason score ≤6) as was a higher PSA (16.6% for each unit increase, 1 ng/mL, in PSA-levels).[172]

The results from **paper IV** confirm that age is an important driver for overdiagnosis. For instance, the cumulative risk of being diagnosed with PC after 5 screens at the age of 60 years was 8%, compared to 13% at 65 years and 21% at 70 years. There are several explanations as to why the risk of overdiagnosis increases with age. As illustrated in autopsy studies, the prevalence of PC increases heavily with age and the pool of latent cancers that can potentially be detected by screening is larger.[20] Another reason as to why age is an important driver of overdiagnosis is that the risk of dying from other competing causes increases with age.[353-355] Therefore, even aggressive cancers can be considered overdiagnosed if diagnosed in older men with limited life-expectancy, who are more likely to die of something else. A study by Albertsen et al. showed that despite having an aggressive cancer (T1c, Gleason score 8-10), a man with comorbidities was up to five times more likely to die from other causes than his PC.[353] A study by Landsdorp-Voogelar et al. used seven different microsimulation models to investigate the benefits and harms of screening and to assess individual cessation ages for breast, colorectal and PC screening based on comorbidity status. The models simulated a U.S cohort of individuals aged 66 to 90 years

who had been screened regularly from 50 years and followed for a lifetime. The model suggested that the "optimal" age for screening cessation, based on comorbid conditions, varied with a 10 year interval around the age cut-point of 74. For example, a person who had been regularly screened since age 50 years, with no comorbid conditions, could continue to be screened until age 74 and still have the same ratio of benefits to harms as someone with average health until age 72 years or as someone with severe comorbid conditions screened until age 66 years.[356] Similar findings were reported from a simulation model using ERSPC data, where screening until the age of 75 in men with low comorbidity had almost the same adjustment for QoL as screening until the age of 69 years in the general population.[170]

Screening factors

Screening factors such as the number of screens, the screening interval, the PSA threshold for biopsy and number of biopsies are also drivers of overdiagnosis. As with age, the association between these risk factors and overdiagnosis has mainly been studied in modeling studies.[19, 165, 170, 171, 173] These studies have found that overdiagnosis is associated with the screening interval, and that a shorter screening interval increases risk of overdiagnosis. However, the screening interval and the PSA threshold appear to have a smaller effect on risk of overdiagnosis than age. De Carvalho et al. investigated 83 different screening policies in a US population with the MISCAN model and compared these policies to a base model defined as annual screening from 50 to 74 years with a PSA threshold of 3 ng/mL. Decreasing the age of screening cessation had a larger effect on reducing overdiagnosis than increasing the screening interval or the PSA threshold.[173] Similar finding was reported from the modeling study based on ERSPC data, mentioned above, where decreasing the cessation age from 74 years to 69 years led to a larger reduction in overdiagnosis than increasing the screening interval from one to four years. Screening annually up to the age of 74 years led to the largest reduction (-32%) in life-years gained due to negative QoL effects, whereas this figure was almost identical between screening annually (-23%) or every four years (-21%) from 55-69 years, implying that annual screening up to the age of 74 had the largest negative effect on QoL.[170]

The risk of overdiagnosis and the screening interval is not straightforward. The screening interval is linked to the number of screens and each screening visit implies a certain risk of overdiagnosis. Moreover, natural fluctuations in the PSA-levels add to the complexity, as the PSA-level fluctuates also for men without prostatic diseases. For example, in the Göteborg screening trial, 17% of men who had an elevated PSA (≥3 ng/mL) but a benign biopsy, had

normalized PSAs during a follow-up of four years.[357] With frequent screens, there is a risk that one of these normal fluctuations results in a positive screening test and subsequent biopsies with detection of an insignificant cancer, that was unrelated to the temporary PSA increase. For example, in the Göteborg screening study, 24% of those with a PSA-level above threshold had PC in subsequent biopsy[119] and of these, 34% were very-low-risk (**paper II**) which is presumed to be clinically insignificant. This implies that among those who had a positive screening test, the risk of being overdiagnosed was 8% (0.24x0.34=0.08).

A long screening interval theoretically exposes the individual to fewer screens over time which would thus reduce overdiagnosis. However, an overly long screening interval may also risk missing the more aggressive, fast-growing tumors and instead diagnose the more slow-growing and less aggressive tumors (length-time bias). This would act in the opposite direction, and result in a larger proportion of less aggressive tumors, that have a higher probability of being overdiagnosed. A shorter screening interval would instead theoretically be associated with a higher likelihood of diagnosing the more fast-growing tumors, which are less likely to be overdiagnosed. Wu et al. investigated the risk of overdiagnosis and the "number of screens for overdetection" (NSO) in the Finnish center of the ERSPC. NSO indicates the expected number of men screened to result in one overdiagnosed man (a lower NSO indicates more overdiagnosis).There was a large difference in NSO depending on the age at starting screening, for example NSO was 104 for men starting at age 55 compared to 44 for starting at age 67, i.e. more overdiagnosis if screening started at a higher age. NSO also decreased with the number of screens and with a shorter screening interval (i.e. more overdiagnosis with more screens and higher screening frequency). However, the effect of the screening interval on risk of overdiagnosis was most pronounced when the screening interval was shortened from 8 to 4 years, NSO was identical for a 1-year and a 2-year screening interval.[165] In **paper IV** we found that the risk of being diagnosed with PC increased with each screen up to four screens but thereafter reached "steady-state". It appeared as if the effect of the prevalent cases was "washed out" after four screens and the number of screens had a smaller effect on the risk of being diagnosed with PC, than age. The results from **paper IV** also suggest that men can start screening at a young age (≈50 years) without substantially increasing their risk of overdiagnosis. This is an important finding as the age at start of screening has been shown to be important for screening to effectively reduce PC mortality.[2] However, the optimal age to start screening has yet to be established.

The PSA threshold is also important for risk of overdiagnosis, but perhaps less so than age and comorbidities.[171, 173] When Gulati et al. modeled the outcome of 35 different screening strategies they found that lowering the PSA threshold from 4 ng/mL to 2.5 ng/mL generated substantial harms in the form of false-positive screening tests and overdiagnosis, relative to the increase in averted deaths from PC.[171] A study by Vickers et al. investigated empirical estimates of overdiagnosis based on age and PSA-level. They found that for men at the age of 60 years the risk of overdiagnosis was relatively low for those with a PSA above 4 ng/mL. The ratio of the risk of clinically detected PC to the risk for biopsy-detected PC for this group of men was close to one. For lower PSA-levels this ratio increased to slightly above two, indication more overdiagnosis.[198]

In **paper IV**, lowering the PSA threshold in the Göteborg screening study was not associated with any apparent change in the cumulative PC incidence for the study population. The PSA threshold was lowered on two occasions (to 2.9 ng/mL in year 1999 and to 2.5 ng/ML in 2005). When analyzing the cumulative PC incidence curve for the entire study population, there was no sharp increase in incidence during the years following the two occasions at which the PSA threshold was lowered; rather, the slope of the curve was relatively constant during the entire follow-up (Figure 11).

The number of biopsy cores sampled at each biopsy affects the detection rate of PC and therefore, most likely, the risk of overdiagnosis. Theoretically, increasing the number of cores should increase the risk of, by chance, detecting very small, potentially insignificant, tumors. During the last 30 years, there has been a trend towards an increasing number of cores, from 2-4 cores in the early 1980's to today's recommendation of 8-12 cores.[11] With the aim of minimizing sampling error some even advocate biopsy schemes including up to 30-40 cores. In the Göteborg screening trial, the number of biopsy cores was increased from 6 to 10 cores in 2009, which potentially could have augmented the risk of overdiagnosis. However, in **paper IV**, we could not identify any discernible increase in the cumulative PC incidence after changing the number of sampled cores (Figure 11). To the best of our knowledge, no study has investigated the number of biopsy cores and risk of overdiagnosis, but several studies have reported that increasing the number of cores detects more insignificant cancers. In the majority of these studies, no statistically significant difference has been observed in the detection of insignificant cancer when comparing sextant biopsies and extended biopsy schemes (10-12 cores), whereas extending the number of cores beyond 12 appears to increase risk of finding insignificant cancer, while having only marginal effect on the overall detection rate.[132]

In summary, population factors such as age and comorbidity appear to be some of the strongest drivers of overdiagnosis. Factors that indicate screening intensity such as the PSA threshold, the number of sampled cores and the screening interval and/or the number of screens probably also affect risk of overdiagnosis, but perhaps not as strongly as the population factors.

Organized versus opportunistic screening

As previously described, screening for PC is a careful balance of benefits and harms. While trying to minimize overdiagnosis, a screening program must simultaneously aim at maximizing the benefits in the form of a reduced morbidity and mortality from advanced PC. In paper III we found that organized, biennial screening was more effective in reducing PC mortality than opportunistic screening, at least in the way it has been performed in the Göteborg male population the last 20 years. The participants in the screening arm of the Göteborg screening study have been screened intensively; every second year with a PSA threshold of 2.5-3 ng/mL, and the vast majority (87%) has complied with the biopsy recommendation following a positive screening test. Unfortunately, we did not have data on exactly how opportunistic screening had been carried out in the Göteborg area during the last 20 years; however, we find it unlikely that it has been as rigorous and intense as the organized screening in the screening arm of the study. For example, in the Dutch center of the ERSPC, only 7-8% of those with a PSA ≥3 ng/mL in the control arm had a biopsy within six months following the PSA-test.[358] The results from **paper III** are in line with several studies from breast- and cervical cancer screening, which have shown that organized screening performs better and is more cost-effective than opportunistic screening.[257, 359, 360] The Council of the European Union has therefore, during several years, recommended that screening for breast, cervical and colorectal cancer should be conducted in organized programs.[361] There is no such recommendation for PC and with the exception of Lithuania,[362] no country in Europe has implemented a national organized PC screening program. The fascinating question is: why is organized screening more effective in reducing PC mortality, and associated with less overdiagnosis (in relation to every averted PC death) than opportunistic screening? What are the differences between these two screening strategies? There are several potential explanations, including screening intensity, the population being screened, compliance with biopsy and time to biopsy following a positive screen, adequacy of the follow-up of a positive screen and different treatment patterns.

The intensity of a screening strategy consists of several different components as previously described (screening interval, PSA-threshold, number of

sampled cores). A comparison between the Dutch center (4 year interval) and the Swedish center (2 year interval) in the ERSPC found that a 2-year interval significantly reduced the incidence of advanced cancers but at the cost of an increased risk of low-risk PC diagnosis, compared to the 4-year interval.[363] These results are supported by the findings of a study from the Finnish ERSPC center, in which the screening interval had a large effect risk of advanced PC and PC mortality. Annual screening reduced risk of advanced PC by 40% compared to 24% for biennial screening and the reduction in PC mortality declined steadily as the screening interval increased.[364] The importance of screening intensity is also supported by a study by Stattin et al. who found that more intense opportunistic screening resulted in lower PC mortality as compared to less intense opportunistic screening. The relative risk of PC mortality in high versus low incidence counties in Sweden, adjusted for time period, was 0.81.[365] A higher screening intensity and over a longer period of time (started earlier in time) is likely also an explanation to why opportunistic screening has been effective in reducing the age-adjusted PC mortality in the US but not in Sweden. Since the introduction of PSA, the age-adjusted PC mortality in the US has fallen by more than 40% from its peak in 1993.[366] Sweden, which has one of the highest PC mortality rates in the world, has had an almost stable PC mortality since the 1960s.[15]

There are no studies that have randomized men to screening with different PSA cut-offs for biopsy but from the PCPT trial it became clear that PSA could not be treated as a dichotomous variable; instead there was a continuum of risk of PC, and PC was present at all PSA-levels. Of these men with a PSA <4 ng/mL and a normal DRE, 15% harbored cancer. However, only 15% of these were Gleason score 7 or higher (corresponding to 2.5% of the entire population) indicating that the majority of tumors detected in these low PSA-levels are low-risk PC with a high risk of being overdiagnosed.[122] Nevertheless, there was no PSA-level where high risk PC was not present and there was no threshold associated with high sensitivity and specificity simultaneously.[121] The Göteborg screening study, which has had the most intense screening algorithm of all the ERSPC centers, with a screening interval of 2 years and a PSA threshold lowered to of 2.5 ng/mL during the course of the trial, is also the center that has reported the largest reduction in PC mortality. These results are a further indication of the importance of screening interval and PSA threshold. The Göteborg screening study is also unique in that it screened younger men, starting at age 50. [2, 179, 180] Another factor that may play a role in the effectiveness of PSA-testing in reducing PC mortality is the time elapsed between a raised PSA to PC diagnosis, and treatment. Although we did not have information

on this, it can be hypothesized that this time period was longer for cases detected with opportunistic screening, which could have led to a worse prognosis and a smaller chance of cure, compared to cases detected with organized screening with a study protocol recommending biopsy for all screen-positive above 2.5-3 ng/mL.

Other possible, theoretical, reasons why opportunistic screening may be less effective than organized screening may be inadequate treatment together with screening of a population who is unlikely to benefit from due to age or comorbidities. However, these factors are unlikely to explain the difference in effectiveness between organized and opportunistic screening in **paper III** because of the reasons discussed below, these factors may be influential when discussing screening in the general population. In the Göteborg screening study, which is a randomized trial, the screening and control arm had the same age distribution. In addition, the majority of men diagnosed with PC, in both the screening and control arm, have been handled by a small group of urologists at the same urology department in Göteborg, where the same treatment strategies apply. Therefore, it is unlikely that differences in age distribution or differential treatment can explain the large difference in the ability to reduce PC mortality. As regards opportunistic screening in the general population, there is room for improvement when it comes to focusing screening efforts on men who can potentially benefit from it. Drazer et al. reported from the National Health Interview Survey during years 2000-2005 in the US that PSA-testing rates were highest for men 70-74 years, among which 45.5% reported being screened in the previous year and screening rates for men 85 years or older was as high as 24.6%. These men are unlikely to reap any benefits from PSA-testing due to competing risks and lower likelihood of being fit for curative treatment at such high age. Although PSA screening rates were lower for men with short life expectancy, 30% of those with the shortest life expectancy (>48% probability of death within 5 years) reported being screened in the previous year.[367] Similar finding have been reported in Sweden from the Stockholm area, were PSA-testing was most common in the age group 70-79 years. In this age group, almost half (45.7%) had undergone PSA-testing within the last two years. In addition, re-testing was common, regardless of the PSA-level. For example, almost half of men 70-79 years who had a PSA <1ng/mL were re-tested within 2 years, despite that their risk of lethal cancer should be very low at such low PSA-levels. [34, 270]

In summary, the results from **paper III** and **paper IV** suggest that screening should ideally be performed with a certain intensity to effectively reduce PC mortality and that opportunistic screening in Göteborg, Sweden during the

last 20 years does not seem to have reached this level. It also suggests that screening can be relatively intense for "younger" (<60-65 years) men, but in order to reduce overdiagnosis screening ought to be performed more selectively in older men and those with comorbidities. Drivers for overdiagnosis are found both in the population screened, as well as in the screening algorithm. By identifying these and by estimating their relative independent and interactive effects, we can create "smarter" screening strategies that minimize the harms while maintaining, or even improving, the benefits.

Methodological considerations paper II and IV

In both **paper III** and **IV** we drew conclusions regarding overdiagnosis based on PC incidence estimates. In **paper III** we assumed that the PC incidence trends in both the screening and control group were largely affected by screening (organized versus opportunistic). The optimal study design would be to have access to data regarding the extent of opportunistic screening in the control group; i.e. the number of PSA-tests, the proportion of PSA-tests that resulted in a biopsy etc. Had we had access to such data, we could have performed a direct comparison of the effects on organized and opportunistic screening on PC mortality, and we could also have evaluated which features of the opportunistic screening that made it less effective in reducing PC mortality. The historical comparison has weaknesses and it is impossible to rule out that there have been changes in the recording, treatment and prognosis of PC. In addition, it is impossible to rule out that there have been no changes in the background risk of PC. Nevertheless, as the Göteborg screening study is a randomized study, any changes would have influenced both arms in similar fashion, and should not have affected the conclusion.

It can be questioned whether NND is the best estimate of overdiagnosis. There is no universally agreed upon method to estimate overdiagnosis. The results from paper III showed that screening, regardless if it is organized or opportunistic resulted in an increased number of men diagnosed with PC compared to a situation without screening. This increase in incidence was smaller for opportunistic screening than for organized screening (Figure 8). Therefore, if overdiagnosis had instead been measured as the excess incidence, opportunistic screening would have appeared as the more favorable alternative. Screening results in four groups of patients; those who are overdiagnosed (not detected in the absence of screening), those who are detected earlier but still too late and will develop symptoms but not die from PC, those who are diagnosed earlier but still too late and would die from PC anyway and those who are saved from PC due to early detection through screening (Figure 12). The only group of these four who benefit from

screening is the last, the other three groups are harmed. The results from **paper III** showed that organized screening was much more effective in reducing PC mortality than opportunistic screening. The primary aim of screening is not to minimize the number of men with the diagnosis but to reduce disease-specific mortality. We therefore believe the group of men whose PC death has been prevented with screening men should be related to those who constitute the extra cases (Figure 12). We therefore choose to estimate overdiagnosis with NND which relates the mortality reduction to the excess incidence.

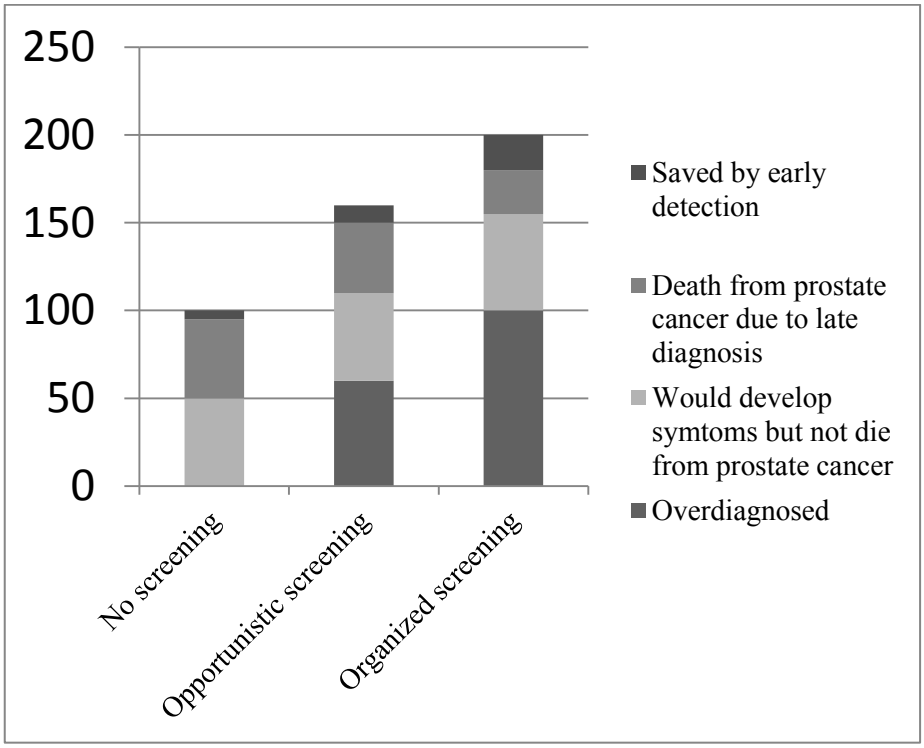

Figure 12. Distribution of patient categories with different methods of diagnosis.

In **paper IV**, as mentioned above, we also used PC incidence to estimate the effect of the number of screens and age at diagnosis on the risk of PC. We assumed that a higher risk of PC diagnosis overall was a reflection of a higher risk of overdiagnosis. Although it can be presumed that an increase in the risk of PC diagnosis reflects an increased risk of overdiagnosis we do not

know how risk of overdiagnosis was distributed among different ages. Furthermore, we were unable to investigate the effect of different screening intervals, but instead estimated the effect of number of screens on incidence. Another limitation is that we were unable to measure the individual effect of each variable (age, number of screens, number of biopsy cores, PSA threshold) on the risk of PC diagnosis. This could possibly be analyzed in the form of a regression analysis, but such an analysis is not straightforward as several of the variables interact and are also time-dependent.

6 GENERAL DISCUSSION AND FUTURE PERSPECTIVES

Shifting the ratio of benefits to harm for PC will require "smarter" screening strategies. While awaiting future discoveries we must make the best out of tools we have available today. There are several strategies that can be applied to screen "smarter" and reduce the harms of overdiagnosis and overtreatment. The aim should be to identify those men who have an elevated risk of developing potentially life-threatening PC. Carlsson and Vickers et al. recently suggested five "golden rules" to improve the benefit to harm ratio for PC screening.[113, 368] To a large extent, these rules summarize what will be discussed below:

- "Get consent"
- "Don't screen men who won't benefit"
- "Don't biopsy without a compelling reason"
- "Don't treat low-risk disease"
- "If you do have to treat, do so at a high-volume center"

6.1 Risk stratified, individualized screening: finding the "right" tumor in the "right" patient

Identifying those with an elevated risk of developing a life-threatening PC and those whom screening should be focused on can be based on age, comorbidities, ethnicity, family history and baseline PSA.

As previously discussed, men of African American origin and men with a family history of PC are at higher risk of PC diagnosis and PC death and are more likely to benefit from screening.[10, 352] For example, in the Swedish National Guidelines for PC care it is recommended that men with hereditary PC (defined as ≥2 close relatives with PC of whom one was diagnosed <75 years of age and/or men who test positive for BRCA2 mutation) start screening at the age 40-50 years and are screened with a 1-2 years interval.[64]

Evidence regarding which age group to screen comes mainly from RCTs restricted to men 50-69 years.[2, 59] In addition, treatment benefits for radical prostatectomy is mainly limited to men <65 years, whereas individual

estimations of benefit apply.[67, 369] The results form **paper IV** and abundant evidence from other studies show that overdiagnosis increases with age.[19, 165, 172, 173, 198] Therefore, there is little evidence for continuing screening above the age of 70 years. On the other hand, there is also evidence indicating that older unscreened men have a higher risk of being diagnosed with higher risk PC with concomitant poorer prognosis, compared to younger men.[370, 371] Nevertheless, in order to minimize overdiagnosis, screening ought to be performed in a more selective manner for men above the age of 70 years. Instead of applying a universal stopping age, an individualized stopping age is probably preferable as comorbidity and remaining life-expectancy can vary greatly.[356] For men above 70 years but with an excellent health, who might have up to 30 years of remaining life-expectancy, screening can still be an alternative. Following the same argument, screening may be less valuable for certain younger men with a high level of comorbidity. Alternative strategies to reduce overdiagnosis in older men would be applying a higher PSA threshold for biopsy and/or increasing screening interval.[171] Several studies from the Malmö Preventive Project and the ERSPC have shown that the PSA-level in midlife is a strong predictor for future risk of clinical PC, PC metastases and PC death [32, 34, 372-375] For instance, one of these studies which compared the screening group in the Göteborg screening study to an unscreened cohort in Malmö, reported that a baseline PSA at age 60 could be useful for identifying those who might benefit from screening. In the 26% of men who had a PSA ≥2 ng/mL continued regular screening led to a large reduction in mortality with a NNS of 23 and NND of 6. The two thirds of men who had a baseline PSA-level at 60 years of < 1ng/mL had very low-risks of PC metastases and death within 15 years.[373] Another study from the same group, showed that the risk of dying from PC at age 85 was very low (≤ 0.2%) for those with a PSA below the median (≤ 1 ng/mL) at age 60.[34] These findings suggests that the baseline PSA-level in midlife can be used to risk-stratify men, such that those with the lowest PSA can be screened less frequently and those with a PSA <1 ng/mL at the age of 60, i.e. the majority of men, can possibly be exempted from further screening.

In order to risk stratify men for screening, nomograms and risk calculators can help find patients at higher risk of PC whom are candidates for biopsy. These instruments are multivariate prediction tools, that have been developed with the aim of reducing unnecessary biopsies and overdiagnosis. Three commonly used risk calculators are the PCPT-risk calculator, the ERSPC risk calculator and the Sunnybrook risk calculator. They are all easily available online.

The PCPT risk calculator is a continuous multivariable risk calculator that was developed based on the 5519 men in the control arm of the trial.[116] The "basic" risk calculator uses race, age, PSA-level, family history, DRE and any previous prostate biopsy to assess the risk of PC if a biopsy is performed.[116] An updated 2.0 version has also been released which can predict low- versus high-risk cancer.[376] The calculator can be used for men above the age of 55 years and who have been screened with PSA and DRE. The calculator has an AUC between 0.56 and 0.72 for predicting a positive biopsy in different cohorts.[377] One limitation with this calculator is that it was based on men of whom the majority had PSA<3 ng/mL and the risk calculator can therefore overestimate the risk of cancer for men with a PSA above this level.

The Dutch arm of the ERSPC has developed a risk calculator in six levels based on six different multivariable logistic regression analyses. Level 1 and 2 are aimed at non-medical persons and do not require any medical knowledge; the remaining levels are for urologists. Level 3 predicts the presence of cancer in sextant biopsy and the degree of aggressiveness for previously unscreened men. Level 4 is for previously screened men but non-biopsied men or those with a prior benign biopsy. Level 5 calculates the chance of harboring indolent cancer. Level 6 predicts future risk of PC.[378] These risk calculators use age, PSA, prostate volume, DRE findings, TRUS result, previous biopsy result and family history. The AUC for level 3-5, which have also been externally validated [379], ranges between 0.68 and 0.79. Using a PSA threshold of 3 ng/mL, sextant biopsy and probability cut-off of 12.5% for a positive biopsy Robool et al. showed that 33% of all biopsies could have been avoided using the risk calculator, compared to a using PSA only.[380] In a head-to-head comparison, the ERSPC risk calculator outperformed the PCPT risk calculator.[381]

The Sunnybrook risk calculator uses age, family history, ethnicity, LUTS, total PSA, F/T PSA ratio and DRE to calculate an individual's risk for PC. The AUC for the nomogram predicting any PC is 0.74 and 0.77 for high-risk cancer.[382] In head-to-head comparisons, the Sunnybrook risk calculator performed better than the PCPT risk calculator but neither one added clinical benefit for risk thresholds of PC diagnosis below 30% [383].

Another possible theoretical strategy to reduce overdiagnosis is primary prevention trough chemoprevention with 5-alfa reductase inhibitors. Finasteride and duasteride have shown to reduce the risk of PC Gleason score ≤6.[384-386] In addition, these agents reduce the symptoms from BPH and decrease the risk of urine retention. Duasteride have shown to improve

sensitivity and specificity of PSA for diagnosing Gleason score 7-10. However, these agents do not reduce the risk of high-risk cancer, or PC death, and concerns have been raised whether finasteride may even increase risk of high-grade cancer.[387] The costs and side-effects should also be considered and the exact role for these agents in the prevention of PC remains to be established.

6.2 Reduce the harms of diagnosis

Today, a large proportion of men with screen-detected PC are treated for a cancer that in the absence of screening would never have been diagnosed. These men have very little to gain from being diagnosed or treated and only risk having to live with side effects from treatment. When Heijnsdijk et al. performed a computer simulation study, modeling the effects of introducing annual screening, the beneficial effects of screening in terms of reducing PC mortality were reduced by 23% of QALYs gained, due to the downstream side effects of overdiagnosis and overtreatment. This can be compared to breast cancer screening where the corresponding figure is a reduction by 8%.[170]

The results from **paper II** in the present thesis, showed that a large proportion of men screen-detected PC are potential candidates for active surveillance. Active surveillance holds many promises for the future and it is, and will continue to be, an important strategy to reduce overtreatment. However, as discussed in chapter 5.2 in this thesis, some challenges need to be faced before active surveillance can be widely accepted. The limited follow-up of the reported active surveillance series and the lack of evidence from RCTs make it questionable to recommend active surveillance to the youngest patients who might have a remaining life-expectancy of more than 30 years. On the other hand, these patients are probably those who have the most to gain when it comes to avoiding or delaying the side-effects from curative treatment. For these men, active surveillance is more about delaying treatment for some years rather than avoiding it completely. Therefore, discontinuing active surveillance and receiving curatively aiming treatment cannot be regarded as failure of active surveillance.

Avoiding overtreatment is as important as avoiding undertreatment in men with potentially life-threatening PC. Akre et al. reported from PCBaSe that only 18% of men with locally advanced PC received curative treatment. PC mortality for those who were not treated was high (8-year PC mortality 28-64%).[388] These figures can be compared to data from SEER-Medicare in the US, where 49% of men with locally advanced PC received curative

treatment within six months from diagnosis.[389] These figures, together with the fact that opportunistic screening in Sweden has not been as intense as in the US, may in part explain why PC mortality in Sweden has been relatively stable since 1960, while in the US, PC mortality has decreased by more than 40%. Curative treatment should be delivered with highest possible quality in men who are likely to benefit and the aim should be to minimize the side-effects from curative treatment, which have profound negative effects on HRQoL.[170] Several studies have shown that oncologic and functional outcome after radical prostatectomy vary between surgeons to a greater extent than what can be explained by chance. For example, Carlsson et al. showed that the individual surgeon's experience was related to the risk of urinary incontinence.[75] A recent review, mostly based on US studies, reported a strong association between hospital volume and outcomes such as perioperative mortality and length of stay and surgeon volume with functional and oncologic outcomes.[390] In Sweden, there has been a centralization of urologic cancer surgery during recent years. For example, in the Western Region (*sv.*"Västra Götalands Regionen"), radical prostatectomy has been centralized to be performed at the three largest hospitals in the region, as compared to being performed at six hospitals earlier. A US study reported that, in 2005, 25% of surgeons performing radical prostatectomy in the US, performed only one procedure per year, and 80% performed less than 10 procedures annually.[391]

6.3 Other strategies to reduce the harms

The results from **paper III** showed that opportunistic screening, the way it has been performed in Göteborg, Sweden during the last 20 years, has increased PC incidence but has had, little, or no effect on PC mortality. Opportunistic screening consumes a fair amount of health care resources, which could be questioned from an ethical perspective. On the other hand, introducing an population-based, organized PSA screening program today seems premature. The levels of overdiagnosis and overtreatment are unacceptable with current screening. "Smarter" screening strategies to reduce the harms are needed before organized screening can be implemented. Nevertheless, the ongoing opportunistic screening needs to be somehow addressed. Possible ways forward could be changing national guidelines together with educational efforts directed at primary care providers and other health care facilities which deliver opportunistic screening. An alternative strategy could be to build specialized screening clinics to which men who are interested in PC screening can be directed. These clinics would then take care of the entire chain of events from PSA-testing, to prostate biopsy to referrals

to urology or oncology clinics. The National Board of Health and Welfare in Sweden currently do not recommend organized PSA screening for men of any age, but state that men aged 50 to 70, who has received written information regarding the pros and cons of PSA screening, and who wish to have a PSA test should have the possibility. The Board also recognizes the need of improving the way current opportunistic PSA screening is carried out and recommend that methods of organizing opportunistic screening, such as standardized routines for delivering PSA and biopsy test results, as well as predetermined screening intervals, should be systematically evaluated. Informed decision-making is a key feature in screening recommendations from a number of organizations. However, this is also an area that can be improved. A study from the NPCR showed that only 14% of Swedish men had received written information before being tested and 10% were unaware that a PSA-test had been ordered.[392] A Swedish study in Region Skåne investigated the feasibility of organized information and voluntary screening by sending out the Board's brochure "About the PSA-test" (*sv.*"Om PSA-prov"), together with a questionnaire and information about the possibility to have a PSA test, to Swedish men aged 50-70. Men were well aware of the possibilities of PSA-testing, but the awareness varied with educational level; from 78% among men with a high-school education to 90% among those with college/university education. There were larger differences regarding the awareness of negative aspects of PSA-testing across educational levels. Of those with college or university education, 72% were aware of the negative aspects of PSA screening, compared to only 42% of those with primary education. These figures suggest that today's Swedish guidelines, under which organized screening is discouraged, but no man can be denied testing if well-informed, risks leading to health care disparities, if well-educated men are those that best can take an informed decision regarding PSA-testing. Whether or not a man is entitled to a PSA-test should not be dependent on the personal beliefs of the treating physician or whether this man is someone who can speak for his rights. Equal health care is a fundamental right in Sweden regulated by the Health care act (*sv.* "Hälso- och sjukvårdslagen"). From this perspective, it might be better with an organized program that can reach the broader population in a more equal manner.

The concept of overdiagnosis has been introduced relatively slowly over time, also among the medical profession. However, there is a growing awareness of this terminology and in 2012, the American National Cancer Institute organized a meeting specifically aimed at evaluate the challange with overdiagnosis in cancer screening. A working group from this meeting constructed a set of recommendations to form a strategy to address the issues surrounding overdiagnosis in cancer screening and prevention. First, it must

be recognized that overdiagnosis occurs, and that it is common. Second, a change in cancer terminology was recommended such that lesions with very low likelihood of progression would not be called “cancer” but instead be labeled “indolent lesions of epithelial origin” (IDLE). For PC, IDLE could be represented by Gleason score 6, according to some authors.[393] This change in taxonomy could help clinicians and patients choosing less aggressive treatment. Third, it was recommended that observational registers be created for low malignant lesions to gain more knowledge regarding these lesions. Fourth, the working group recommended that overdiagnosis should be mitigated with strategies such as avoiding diagnostic assessment when it is not truly necessary, reducing the screening frequency, raising the threshold for what constitutes a positive screen, and focusing screening on high-risk individuals. Fifth and lastly, they recommended that new strategies for determining what defines cancer progression and new screening strategies should be embraced.[394]

There is an urgent need to improve the screening tools and diagnostic tools for PC. We are continuously gaining new knowledge and the future will hopefully bring us better tools such as biomarkers, genetic tests and novel imaging techniques that can help us identify and treat only those men who will actually benefit. Until then, we can do a lot by using the currently available tools in a “smarter” way.

6.3.1 Novel screening markers, biomarkers

The optimal biomarker for screening should be non-invasive, non-expensive, specific for PC and be able to differentiate between aggressive and non-aggressive PC. However, this marker is yet to be found.

Work is ongoing aiming at further improving the widely used PSA assays. The four kallikrein panel (4K score) [395] and the prostate heath index (PHI) are amongst the biomarkers that are the furthest ahead but both need additional evaluation.[396] The four kallikrein panel (total PSA, free PSA, intact PSA, and hK2) appear to have an improved predictive accuracy for diagnosing PC and high risk PC and have the potential to reduce the number of biopsies.[395] PHI combines total PSA, free PSA and -2proPSA into a single score and is calculated as PHI=[-2proPSA/free PSA] x √PSA. Catalona et al. reported, from a prospective multicenter study, that for men with a PSA 2-10 ng/mL, at a sensitivity of 80-95%, PHI performed better (specificity and AUC) than PSA and free/total PSA alone.[397] There are also several urinary markers based on RNA, DNA or proteins under investigation. The RNA-based urine markers such as the Prostate cancer

antigen 3 gene (PCA3) and TMPRSS2:ERG are amongst the most developed. However, urinary tests are less feasible than blood tests, since they must be preceded by prostatic massage by DRE to obtain enough cells to be analyzed. PCA3 is a prostate-specific gene and PCA3 mRNA is overexpressed in many PC cells. PCA3 score measures the amount of non-coding PCA3 mRNA in relation to normal PSA mRNA. Currently, the main use of PCA3 is in determining the need of repeat biopsy in men with an initial negative biopsy.[11] Detection of the gene fusion of TMPRSS2:ERG in urine may have the potential to better distinguish low-risk tumors from more aggressive ones.[398]

Genetic and epigenetic markers of PC risk and prognosis is also an area of intense research. Genome-wide association studies (GWAS) have tried to identify markers of increased PC risk but the genetics behind PC is heterogeneous and so far, no single gene has been identified. Certain single nucleotide polymorphisms (SNP) have been linked to PC risk but the change in risk associated with each SNP is small. However, risk of PC appears to increase with increasing number of risk alleles.[399] Epigenetic changes in PCs are also a promising area of research. DNA methylation, for example, can lead to silencing or amplification of genes resulting in gene-expression alteration without changing the DNA sequence. For example, methylation status of the GSTP1 gene may be associated with PC risk.[400]

6.3.2 Magnetic Resonance Imaging

Multiparametric Magnetic Resonance Imaging (mp-MRI) has emerged as a promising adjunct to PSA in the diagnosis and follow-up of PC. MRI has the advantages of being a non-invasive procedure that does not expose the individual to ionizing radiation and can image the entire prostate in 3-5 mm sections. It builds on the principle that atomic nuclei in a strong magnetic field absorb pulses of radiofrequency energy and emit them as radio waves. These radio waves can then be received and reconstructed into 3D-images. [401] mpMRI includes T2-weighted imaging (T2WI), diffusion weighted imaging (DWI) and dynamic contrast enhanced imaging (DCEI). Some centers also perform magnetic resonance spectroscopy (MRS), which measures concentrations of metabolites (choline and creatine) in suspicious areas.

T2WI can define the prostate anatomy with its different zones. Areas of low signal can, especially in the peripheral zone, indicate cancer. DWI measures the diffusion of water molecules. Areas of PC exhibit reduced diffusion of water. With this information, an apparent diffusion coefficient (APC) map of the prostate is built. In the contrast enhanced series a bolus of gandolinium

contrast is given followed by many rapid scans at short intervals. With this information it is possible to calculate time-against-perfusion curves.[401] It has been suggested that, in these curves, high-grade PC exhibit early and intense contrast enhancement and a rapid wash-out, in contrast to lower grade PC

The European Society of Urogenital Radiology (ESUR) recommends that findings on prostate MRI are reported in a structure manner according to the Prostate Imaging-Reporting and Data System (PI-RAD). A score is given for each parameter within each region of interest. Presence of clinically significant cancer is indicated as 1=extremely unlikely, 2=unlikely and so on up to 5=extremely likely.[401, 402] T2WI, DWI and MRS can be helpful in estimating tumor grade and volume. These abilities could help reducing overdiagnosis by more selectively choosing candidates for biopsy and AS and for following men on AS.[401] mpMRI also has the potential to improve the way prostate biopsies are performed. Today, there are at least three main techniques to perform MRI-guided biopsies[401, 403, 404]:

1. "MRI-informed free-hand" (cognitive) technique, during which the operator reviews the MRI images, and with this information performs a TRUS-guided biopsy directed towards the MRI-suspicious area. This technique has the advantages of being cheap and not requiring any specific equipment other than TRUS, but is limited by the obvious difficulties in guaranteeing that the "right" area is sampled.
2. "MRI-TRUS-fusion biopsy", where a particular software program downloads the information from the MRI and transfers it over to the ultrasound machine, and a fused image is built. Biopsies can then be directed towards the MRI-suspicious area. The advantages include a short learning curve for the examiner and possibly more accurate sampling. However, today's equipment cannot compensate for any potential movement and deformation of the prostate by the probe and biopsy gun since the fusion between the MRI and ultrasound images cannot be performed in real time.
3. "Real-time MRI-guided biopsy". This technique is the most advanced and includes a MRI-compatible biopsy device, which allows biopsy sampling within the MRI machine, under the guidance of real-time MRI images. This technique can guarantee sampling from the area of interest, but it

limited by being time-consuming, costly and requiring the need for anesthesia

Standard TRUS-guided biopsies (10-14 cores) are limited by sampling errors, as previously described, with both undersampling (false negatives or missing the foci with the highest Gleason score) and oversampling (diagnosis of clinically insignificant cancer). Undersampling is especially common in the anterior parts of the prostate. MRI-guided targeted biopsies have the potential to address several of these limitations. Among men with no previous biopsy, MRI may increase cancer detection with 50% for low-risk men and 70% for high-risk men. For low-risk men with previously negative biopsies, a negative MRI has a NPV which reaches 98%, implying that biopsies can be avoided and overdiagnosis decreased. For men with a previous negative biopsy, 72-87% of tumors detected under MRI-guidance are clinically significant.[403]

mp-MRI also seems promising for the selection and follow-up of men on active surveillance. Turkbey et al. investigated the ability of a 3 Tesla (a measure of field strenght) mP-MRI and different clinical inclusion criteria (D'Amico, Epstein, CAPRA) to select appropriate candidates for active surveillance by comparing the results with findings at radical prostatectomy. mp-MRI outperformed clinical criteria and had a sensitivity of 93%, a PPV of 57% and an overall accuracy of 92% for finding clinically significant cancer, defined as tumor volume <0.5 ml, no Gleason pattern 4 or 5 or extracapsular or seminal vesicle involvement.[405] Stamatakis et al. reported on 85 patients who met the Johns Hopkins active surveillance criteria and who subsequently underwent a mp-MRI. After a confirmatory MRI/TRUS-fusion-biopsy, 29% were no longer candidates for active surveillance.[406] This figure is similar to the figures of upgrading from standard TRUS-guided biopsies and radical prostatectomy. In a study of 388 men with low-risk PC who underwent MRI and a confirmatory biopsy, Vargas et al. found that findings on MRI correlated well with findings on confirmatory biopsy with a high NPV (86-100%), specificity (95-100%) and sensitivity (87-98%).[407]

These and other studies, indicate that MRI has a high NPV for intermediate outcomes (upgrading) for men on AS and that a high-quality MRI can be of value for patient selection and follow-up to reduce the need for repeat biopsy. Much work remains before we know what constitutes significant radiological disease and radiological progression for men on active surveillance.

The GÖTEBORG-2 trial

Several of these, new, promising features of mp-MRI will be investigated in the *GÖTEBORG-2 trial* which will be launched in late 2014/early 2015. The trial is a 3-arm RCT, which will be a collaboration between the Depertment of Urology and the Department of Radiology at the Sahlgrenska Academy, the Regional Cancer Center and Chalmers University of Technology, in Göteborg, Sweden. In the first phase, 40000 men in the Western Region will be randomized in a 1:1 ratio to a screening or control group. The 20,000 men in the screening group will, after informed consent, be offered a PSA-test. Those who have had a PSA-test will then be randomized into one of three study arms.

- Arm 1 (the reference arm). In this arm, the PSA threshold is 3 ng/mL. Screen-positive men will be invited to a mpMRI followed by TRUS-guided standard (10-core) biopsy for all men, and for those with a suspicious lesion on MRI, an extra 4 cores directed towards the suspicious area/s (targeted biopsies).
- Arm 2 (experimental arm I). In this arm, the PSA threshold is again 3 ng/mL. Screen-positive men will be invited to mpMRI, but only targeted biopsies, and no systematic biopsies, will be performed. Men with a negative MRI will not be biopsied.
- Arm 3 (experimental arm II). In this arm, the PSA threshold is 1.8 ng/mL, and otherwise identical to arm 2.

The hypothesis is that PSA+mp-MRI will reduce the number of men biopsied unnecessarily, and increase specificity and reduce overdiagnosis by reducing the number of men diagnosed with insignificant cancer, without compromising sensitivity or detection rate.

Several side-studies will also be performed, including studies on technical aspects of mp-MRI, feasibility and logistics, costs and cost-effectiveness, QoL, biomarkers, equitable care and health care disparities.

7 CONCLUSION

This thesis aimed at exploring different aspects of overdiagnosis in screening for PC. In **paper I**, we concluded that death certificates for men with PC in Sweden are of high quality and can be used for endpoint evaluation of PC screening studies. Overdiagnosis in the screening arm was not associated with any larger effect on the COD determination. In **paper III** and **paper IV** we investigated drivers of overdiagnosis. By organizing screening within the frameworks of a program the effectiveness, in terms of reducing PC mortality, can be increased and overdiagnosis reduced (relative to the gain).

If a screening program were to be introduced, our results indicate that, in order to minimize overdiagnosis, screening should be focused on younger men and only performed carefully in selected older men. The intensity of screening appears important for screening to effectively reduce PC mortality, but seem to be of less important for the risk of overdiagnosis than age at termination of screening. In **paper II**, we explored how overtreatment, following overdiagnosis, can be reduced. Active surveillance appears to be a promising management strategy, at least until we have a screening strategy that can selectively diagnose only clinically relevant tumors. A large proportion of men diagnosed with screen-detected PC are candidates for active surveillance. With active surveillance these men can avoid or postpone the side-effects of curative treatment without risking the chance of cure, at least in the medium-term.

We now have evidence that screening with PSA can reduce the burden of PC in terms of reduced morbidity and PC mortality [2, 59, 148, 149] but is associated with considerable harms, of which overdiagnosis and overtreatment are the major concerns. Does this mean we should stop the use of PSA as a screening test? The obvious answer for me personally is: “No”. By no means would we would want to turn back time to the pre-PSA era and most would probably agree that it would not be desirable. So, what options do we have? Until alternative screening tools or strategies are developed we must stop using a “one-size fits all” strategy for screening and instead develop individualized screening strategies where focus is on the individual patient.[408] Guided by factors such as the PSA-value, age, comorbidities and life expectancy, we can screen “smarter” and reduce overdiagnosis. Hopefully, in the near future, MRI will be a valuable and integrated part of the screening algorithm that may further reduce the harms of screening.

ACKNOWLEDGEMENT

The work behind this thesis was conducted at the Urology Department at the Sahlgrenska Academy at University of Göteborg between 2009 and 2014. Many people have been involved in this work and I would like to especially express my gratitude to the following people:

Jonas Hugosson, M.D., Professor of Urology at the Sahlgrenska Academy at University of Göteborg, my main supervisor, who introduced me to the world of research. Thank you for believing in me and for giving me this fantastic opportunity to conduct my PhD work within the Göteborg screening study. Although there have been times when finalizing this thesis have felt impossible, I have always left our meetings filled with energy and optimism. You are a true inspiration!

Johan Stranne, M.D., Associate Professor at the Sahlgrenska Academy at University of Göteborg, and my co-supervisor. Thank you for your enthusiasm and for always making me feel welcome; no matter how overloaded with work you are, you always take time to help and assist in every possible way.

Jan-Erik Damber, M.D., Professor at the Sahlgrenska Academy at University of Göteborg Thank you for giving me the possibility to be a PhD student at your department.

Ali Khatami, M.D. PhD, head of the Department of Urology at the Sahlgrenska University Hospital and co-author of paper II. Thank you for hiring me as a junior doctor, fresh out of medical school, for introducing me to the wonderful world of urology and giving me the opportunity to conduct my research at your clinic.

Erik Holmberg, PhD, statistician. Thank you for your fantastic help with many of the statistical analysis of this thesis, for being patient with short deadlines and the numerous e-mails from me even late at night, during week-ends and holidays.

Sigrid Carlsson, M.D., PhD, MPH, Associate Professor the the Sahlgrenska Academy at University of Göteborg. Thank you for being my unofficial mentor during these years, for helping me with everything from revising manuscript (and this thesis) to giving advice on dress codes for various research events.

Helén Ahlgren, database manager. Thank you for your kind assistance with the study database and for all the fun we have had during our work. Without you, this thesis would not have been possible.

Maria Nyberg, research nurse and my roommate during the last six months. Thank you for all your help and patience. Being roommate with a "klok gumma" has been invaluable!

Lars Rex, M.D. and **Hans Boman**, M.D., PhD former and present head of the Surgery Department at Södra Älvsborgs sjukhus. Thank you for enabling this thesis by giving me time off from clinical work to conduct my PhD work.

Hans Lilja, M.D., Professor of Clinical Biochemistry. Thank you for invaluable work with the Göteborg screening study and for co-authoring of paper III.

The members of the cause of death committee of the Göteborg screening study; **Silas Pettersson**, Professor Emeritus in Urology, **Eberhard Varenhorst**, Professor Emeritus in Urology, **Per Folmerz**, specialist in urology and the late **Bo Johan Norlén**, Professor Emeritus in Urology. Thank you for your hard work and for charing your enormous knowledge and experience.

Anna Grenbo Bergdahl, Maria Frånlund and **Kimia Kohestani** my fellow PhD students. Thank you for all the fun we have had in our research team and for the numerous trips around Europe. **Anna**, I owe you a trip to Italy!

All my **Colleagues** at the Urology Department at Sahlgrenska University Hospital. Thank you for always making me feel welcome and for generously charing your knowledge and experience.

To my **mom**, my **dad** and my **brother Markus**, thank you for your unconditional love and for your support, both emotionally and financially. Thank you for always believing in me, yet never putting pressure on me.

To **Martin**, the love of my life, my fantastic husband. Thank you for your patience, love and support and for encouraging me to do thing for the right reasons. To **Josefina**, my beautiful, intelligent daughter, thank you for every day reminding me what is truly important.

Thank you!

This thesis was made possible by the financial support by:

- Swedish Cancer Society (Cancerfonden)
- Percy Falk's foundation for prostate and breast cancer research
- Märta and Gustaf Ågren's foundation
- Anna-Lisa and Bror Björnsson's foundation

REFERENCES

1. Dorlands Medical Dictionary; Available from: http://www.dorlands.com/.[Accessed 6 Oct 2014]
2. Hugosson J, Carlsson S, Aus G, et al. Mortality results from the Goteborg randomised population-based prostate-cancer screening trial. Lancet Oncol. 2010;11(8):725-32.
3. Walsh CP, Retik BA, Vaughan DE, Wein A. Campbell's Urology. 7th ed. Philadelphia: W.B. Saunders Company; 1998.
4. Balk SP, Ko YJ, Bubley GJ. Biology of prostate-specific antigen. J Clin Oncol. 2003;21(2):383-91.
5. The National Board of Health and Welfare S. Cancer incidence in Sweden 2012. 2013; Available from: http://www.socialstyrelsen.se/publikationer2013/2013-12-17.[Accessed 6 Oct 2014]
6. Leitzmann MF, Rohrmann S. Risk factors for the onset of prostatic cancer: age, location, and behavioral correlates. Clinical epidemiology. 2012;4:1-11.
7. National Prostate Cancer Register. Annual report 2012. 2013; Available from: http://npcr.se/wp-content/uploads/2013/04/20131121-NPCR-Rapport-2012.pdf.
8. Sakr WA, Grignon DJ, Crissman JD, et al. High grade prostatic intraepithelial neoplasia (HGPIN) and prostatic adenocarcinoma between the ages of 20-69: an autopsy study of 249 cases. In Vivo. 1994;8(3):439-43.
9. Soos G, Tsakiris I, Szanto J, Turzo C, Haas PG, Dezso B. The prevalence of prostate carcinoma and its precursor in Hungary: an autopsy study. Eur Urol. 2005;48(5):739-44.
10. Hemminki K. Familial risk and familial survival in prostate cancer. World J Urol. 2012;30(2):143-8.
11. Mottet N, Bastian PJ, Bellmunt J, et al. EAU Guidelines on Prostate Cancer. European Association of Urology. 2014.
12. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. 2010;127(12):2893-917.
13. Center MM, Jemal A, Lortet-Tieulent J, et al. International variation in prostate cancer incidence and mortality rates. Eur Urol. 2012;61(6):1079-92.
14. Cancer incidence in Sweden 2011. The National Board of Health and Welfare. Available from: http://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/18919/2012-12-19.pdf. [Accessed 6 Oct 2014]

15. Engholm G, Ferlay J, Christensen N, et al. NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries. Association of the Nordic Cancer Registries. Danish Cancer Society.; [cited 2013 18-11-2013]; Version 5.3 (25.04.2013):[Available from: http://www.ancr.nu.
16. Etzioni R, Tsodikov A, Mariotto A, et al. Quantifying the role of PSA screening in the US prostate cancer mortality decline. Cancer Causes Control. 2008;19(2):175-81.
17. Etzioni R, Gulati R, Tsodikov A, et al. The prostate cancer conundrum revisited: treatment changes and prostate cancer mortality declines. Cancer. 2012;118(23):5955-63.
18. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. CA: a cancer journal for clinicians. 2014;64(1):9-29.
19. Draisma G, Boer R, Otto SJ, et al. Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst. 2003;95(12):868-78.
20. Leal J, Hamdy F, Wolstenholme J. Estimating age and ethnic variation in the histological prevalence of prostate cancer to inform the impact of screening policies. International journal of urology : official journal of the Japanese Urological Association. 2014;21(8):786-92.
21. Konety BR, Bird VY, Deorah S, Dahmoush L. Comparison of the incidence of latent prostate cancer detected at autopsy before and after the prostate specific antigen era. J Urol. 2005;174(5):1785-8; discussion 8.
22. Chodak GW, Thisted RA, Gerber GS, et al. Results of conservative management of clinically localized prostate cancer. N Engl J Med. 1994;330(4):242-8.
23. Johansson JE, Holmberg L, Johansson S, Bergstrom R, Adami HO. Fifteen-year survival in prostate cancer. A prospective, population-based study in Sweden. JAMA. 1997;277(6):467-71.
24. Albertsen PC, Hanley JA, Gleason DF, Barry MJ. Competing risk analysis of men aged 55 to 74 years at diagnosis managed conservatively for clinically localized prostate cancer. JAMA. 1998;280(11):975-80.
25. Aus G, Hugosson J, Norlen L. Long-term survival and mortality in prostate cancer treated with noncurative intent. J Urol. 1995;154(2 Pt 1):460-5.
26. Albertsen PC, Hanley JA, Fine J. 20-year outcomes following conservative management of clinically localized prostate cancer. JAMA. 2005;293(17):2095-101.
27. Cooperberg MR, Broering JM, Carroll PR. Time trends and local variation in primary treatment of localized prostate cancer. J Clin Oncol. 2010;28(7):1117-23.
28. Lu-Yao GL, Albertsen PC, Moore DF, et al. Outcomes of localized prostate cancer following conservative management. JAMA. 2009;302(11):1202-9.

29. Gulati R, Wever EM, Tsodikov A, et al. What if I don't treat my PSA-detected prostate cancer? Answers from three natural history models. Cancer Epidemiol Biomarkers Prev. 2011;20(5):740-50.
30. Hugosson J, Aus G, Becker C, et al. Would prostate cancer detected by screening with prostate-specific antigen develop into clinical cancer if left undiagnosed? A comparison of two population-based studies in Sweden. BJU Int. 2000;85(9):1078-84.
31. Gann PH, Hennekens CH, Stampfer MJ. A prospective evaluation of plasma prostate-specific antigen for detection of prostatic cancer. JAMA. 1995;273(4):289-94.
32. Lilja H, Cronin AM, Dahlin A, et al. Prediction of significant prostate cancer diagnosed 20 to 30 years later with a single measure of prostate-specific antigen at or before age 50. Cancer. 2011;117(6):1210-9.
33. Vickers AJ, Ulmert D, Sjoberg DD, et al. Strategy for detection of prostate cancer based on relation between prostate specific antigen at age 40-55 and long term risk of metastasis: case-control study. BMJ. 2013;346:f2023.
34. Vickers AJ, Cronin AM, Bjork T, et al. Prostate specific antigen concentration at age 60 and death or metastasis from prostate cancer: case-control study. BMJ. 2010;341:c4521.
35. Gleason DF. Classification of prostatic carcinomas. Cancer chemotherapy reports Part 1. 1966;50(3):125-8.
36. Epstein JI, Allsbrook WC, Jr., Amin MB, Egevad LL. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. The American journal of surgical pathology. 2005;29(9):1228-42.
37. Albertsen PC, Hanley JA, Barrows GH, et al. Prostate cancer and the Will Rogers phenomenon. J Natl Cancer Inst. 2005;97(17):1248-53.
38. Helpap B, Egevad L. The significance of modified Gleason grading of prostatic carcinoma in biopsy and radical prostatectomy specimens. Virchows Archiv : an international journal of pathology. 2006;449(6):622-7.
39. AstraZeneca Oncology; Available from: https://www.astrazenecaoncology.se/Tools/Image-bank/.[Accessed 25 Oct 2014]
40. Allsbrook WC, Jr., Mangold KA, Johnson MH, Lane RB, Lane CG, Epstein JI. Interobserver reproducibility of Gleason grading of prostatic carcinoma: general pathologist. Human pathology. 2001;32(1):81-8.
41. Glaessgen A, Hamberg H, Pihl CG, Sundelin B, Nilsson B, Egevad L. Interobserver reproducibility of modified Gleason score in radical prostatectomy specimens. Virchows Archiv : an international journal of pathology. 2004;445(1):17-21.
42. Netto GJ, Eisenberger M, Epstein JI. Interobserver variability in histologic evaluation of radical prostatectomy between central and local

pathologists: findings of TAX 3501 multinational clinical trial. Urology. 2011;77(5):1155-60.
43. Van der Kwast TH, Lopes C, Martikainen PM, et al. Report of the Pathology Committee: false-positive and false-negative diagnoses of prostate cancer. BJU Int. 2003;92 Suppl 2:62-5.
44. Egevad L, Granfors T, Karlberg L, Bergh A, Stattin P. Prognostic value of the Gleason score in prostate cancer. BJU Int. 2002;89(6):538-42.
45. Partin AW, Mangold LA, Lamm DM, Walsh PC, Epstein JI, Pearson JD. Contemporary update of prostate cancer staging nomograms (Partin Tables) for the new millennium. Urology. 2001;58(6):843-8.
46. Briganti A, Gallina A, Suardi N, et al. A nomogram is more accurate than a regression tree in predicting lymph node invasion in prostate cancer. BJU Int. 2008;101(5):556-60.
47. Eifler JB, Feng Z, Lin BM, et al. An updated prostate cancer staging nomogram (Partin tables) based on cases from 2006 to 2011. BJU Int. 2013;111(1):22-9.
48. The National Board of Health and Welfare. Nationella riktlinjer för prostatacancer. Vetenskapligt underlag. Bilaga.2013. Available from: http://www.socialstyrelsen.se/SiteCollectionDocuments/nr-cancer-vetenskapligt-underlag-prostatacancer.pdf [Accessed 6 Oct 2014].
49. Heidenreich A, Bastian PJ, Bellmunt J, et al. EAU guidelines on prostate cancer. part 1: screening, diagnosis, and local treatment with curative intent-update 2013. Eur Urol. 2014;65(1):124-37.
50. Sobin LH, Gospodarowicz M, Wittekind C. TNM classification of malignant tumors. UICC International Union Against Cancer. 7th ed: Wiley-Blackwell; 2009.
51. Whitmore WF, Jr. Natural history and staging of prostate cancer. The Urologic clinics of North America. 1984;11(2):205-20.
52. Briganti A, Passoni N, Ferrari M, et al. When to perform bone scan in patients with newly diagnosed prostate cancer: external validation of the currently available guidelines and proposal of a novel risk stratification tool. Eur Urol. 2010;57(4):551-8.
53. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. Journal of chronic diseases. 1987;40(5):373-83.
54. Rider JR, Sandin F, Andren O, Wiklund P, Hugosson J, Stattin P. Long-term outcomes among noncuratively treated men according to prostate cancer risk category in a nationwide, population-based study. Eur Urol. 2013;63(1):88-96.
55. D'Amico AV, Whittington R, Malkowicz SB, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. JAMA. 1998;280(11):969-74.

56. Cooperberg MR, Pasta DJ, Elkin EP, et al. The University of California, San Francisco Cancer of the Prostate Risk Assessment score: a straightforward and reliable preoperative predictor of disease recurrence after radical prostatectomy. J Urol. 2005;173(6):1938-42.
57. National Comprehensive Cancer Network (NCCN) clinical practice guidelines in Oncology. Prostate Cancer, version I.2014. NCCN.org.
58. National Prostate Cancer Register. Annual report 2013. 2014; Available from: http://npcr.se/rapporter/nationella-arsrapporter/. [Accessed 25 Oct 2014]
59. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Lancet. 2014.
60. Godtman RA, Holmberg E, Khatami A, Stranne J, Hugosson J. Outcome following active surveillance of men with screen-detected prostate cancer. Results from the Goteborg randomised population-based prostate cancer screening trial. Eur Urol. 2013;63(1):101-7.
61. D'Amico AV, Moul J, Carroll PR, Sun L, Lubeck D, Chen MH. Cancer-specific mortality after surgery or radiation for patients with clinically localized prostate cancer managed during the prostate-specific antigen era. J Clin Oncol. 2003;21(11):2163-72.
62. Boorjian SA, Karnes RJ, Rangel LJ, Bergstralh EJ, Blute ML. Mayo Clinic validation of the D'amico risk group classification for predicting survival following radical prostatectomy. J Urol. 2008;179(4):1354-60; discussion 60-1.
63. moHernandez DJ, Nielsen ME, Han M, Partin AW. Contemporary evaluation of the D'amico risk classification of prostate cancer. Urology. 2007;70(5):931-5.
64. Nationellt vårdprogram för Prostatacancer. 2014. Available from http://www.cancercentrum.se/vardprogram/prostatacancer/ [Accessed 9 Nov 2014].
65. Briganti A, Chun FK, Salonia A, et al. Complications and other surgical outcomes associated with extended pelvic lymphadenectomy in men with localized prostate cancer. Eur Urol. 2006;50(5):1006-13.
66. Van Hemelrijck M, Garmo H, Holmberg L, et al. Thromboembolic events following surgery for prostate cancer. Eur Urol. 2013;63(2):354-63.
67. Bill-Axelson A, Holmberg L, Ruutu M, et al. Radical prostatectomy versus watchful waiting in early prostate cancer. N Engl J Med. 2011;364(18):1708-17.
68. Wilt TJ, Brawer MK, Jones KM, et al. Radical prostatectomy versus observation for localized prostate cancer. N Engl J Med. 2012;367(3):203-13.

69. Montorsi F, Wilson TG, Rosen RC, et al. Best practices in robot-assisted radical prostatectomy: recommendations of the Pasadena Consensus Panel. Eur Urol. 2012;62(3):368-81.
70. Novara G, Ficarra V, Mocellin S, et al. Systematic review and meta-analysis of studies reporting oncologic outcome after robot-assisted radical prostatectomy. Eur Urol. 2012;62(3):382-404.
71. Novara G, Ficarra V, Rosen RC, et al. Systematic review and meta-analysis of perioperative outcomes and complications after robot-assisted radical prostatectomy. Eur Urol. 2012;62(3):431-52.
72. Ficarra V, Novara G, Ahlering TE, et al. Systematic review and meta-analysis of studies reporting potency rates after robot-assisted radical prostatectomy. Eur Urol. 2012;62(3):418-30.
73. Ficarra V, Novara G, Rosen RC, et al. Systematic review and meta-analysis of studies reporting urinary continence recovery after robot-assisted radical prostatectomy. Eur Urol. 2012;62(3):405-17.
74. Coelho RF, Rocco B, Patel MB, et al. Retropubic, laparoscopic, and robot-assisted radical prostatectomy: a critical review of outcomes reported by high-volume centers. Journal of endourology / Endourological Society. 2010;24(12):2003-15.
75. Carlsson S, Berglund A, Sjoberg D, et al. Effects of surgeon variability on oncologic and functional outcomes in a population-based setting. BMC urology. 2014;14:25.
76. Vickers A, Savage C, Bianco F, et al. Cancer control and functional outcomes after radical prostatectomy as markers of surgical quality: analysis of heterogeneity between surgeons at a single cancer center. Eur Urol. 2011;59(3):317-22.
77. Bianco FJ, Jr., Riedel ER, Begg CB, Kattan MW, Scardino PT. Variations among high volume surgeons in the rate of complications after radical prostatectomy: further evidence that technique matters. J Urol. 2005;173(6):2099-103.
78. Zietman AL, Bae K, Slater JD, et al. Randomized trial comparing conventional-dose with high-dose conformal radiation therapy in early-stage adenocarcinoma of the prostate: long-term results from proton radiation oncology group/american college of radiology 95-09. J Clin Oncol. 2010;28(7):1106-11.
79. Beckendorf V, Guerif S, Le Prise E, et al. 70 Gy versus 80 Gy in localized prostate cancer: 5-year results of GETUG 06 randomized trial. Int J Radiat Oncol Biol Phys. 2011;80(4):1056-63.
80. Peinemann F, Grouven U, Bartel C, et al. Permanent interstitial low-dose-rate brachytherapy for patients with localised prostate cancer: a systematic review of randomised and nonrandomised controlled clinical trials. Eur Urol. 2011;60(5):881-93.
81. Kupelian PA, Potters L, Khuntia D, et al. Radical prostatectomy, external beam radiotherapy <72 Gy, external beam radiotherapy > or =72 Gy, permanent seed implantation, or combined

seeds/external beam radiotherapy for stage T1-T2 prostate cancer. Int J Radiat Oncol Biol Phys. 2004;58(1):25-33.
82. Sylvester JE, Grimm PD, Blasko JC, et al. 15-Year biochemical relapse free survival in clinical Stage T1-T3 prostate cancer following combined external beam radiotherapy and brachytherapy; Seattle experience. Int J Radiat Oncol Biol Phys. 2007;67(1):57-64.
83. Pommier P, Chabaud S, Lagrange JL, et al. Is there a role for pelvic irradiation in localized prostate adenocarcinoma? Preliminary results of GETUG-01. J Clin Oncol. 2007;25(34):5366-73.
84. Widmark A, Klepp O, Solberg A, et al. Endocrine treatment, with or without radiotherapy, in locally advanced prostate cancer (SPCG-7/SFUO-3): an open randomised phase III trial. Lancet. 2009;373(9660):301-8.
85. Jones CU, Hunt D, McGowan DG, et al. Radiotherapy and short-term androgen deprivation for localized prostate cancer. N Engl J Med. 2011;365(2):107-18.
86. Kumar S, Shelley M, Harrison C, Coles B, Wilt TJ, Mason MD. Neo-adjuvant and adjuvant hormone therapy for localised and locally advanced prostate cancer. Cochrane Database Syst Rev. 2006(4):CD006019.
87. Bolla M, Van Tienhoven G, Warde P, et al. External irradiation with or without long-term androgen suppression for prostate cancer with high metastatic risk: 10-year results of an EORTC randomised study. Lancet Oncol. 2010;11(11):1066-73.
88. Caras RJ, Sterbis JR. Prostate cancer nomograms: a review of their use in cancer detection and treatment. Current urology reports. 2014;15(3):391.
89. Stattin P, Holmberg E, Johansson JE, Holmberg L, Adolfsson J, Hugosson J. Outcomes in localized prostate cancer: National Prostate Cancer Register of Sweden follow-up study. J Natl Cancer Inst. 2010;102(13):950-8.
90. Hoffman RM, Koyama T, Fan KH, et al. Mortality after radical prostatectomy or external beam radiotherapy for localized prostate cancer. J Natl Cancer Inst. 2013;105(10):711-8.
91. Sooriakumaran P, Nyberg T, Akre O, et al. Comparative effectiveness of radical prostatectomy and radiotherapy in prostate cancer: observational study of mortality outcomes. BMJ. 2014;348:g1502.
92. Grimm P, Billiet I, Bostwick D, et al. Comparative analysis of prostate-specific antigen free survival outcomes for patients with low, intermediate and high risk prostate cancer treatment by radical therapy. Results from the Prostate Cancer Results Study Group. BJU Int. 2012;109 Suppl 1:22-9.
93. Rao AR, Motiwala HG, Karim OM. The discovery of prostate-specific antigen. BJU Int. 2008;101(1):5-10.

94. Stamey TA, Yang N, Hay AR, McNeal JE, Freiha FS, Redwine E. Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. N Engl J Med. 1987;317(15):909-16.
95. Catalona WJ, Smith DS, Ratliff TL, et al. Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. N Engl J Med. 1991;324(17):1156-61.
96. Labrie F, Dupont A, Suburu R, et al. Serum prostate specific antigen as pre-screening test for prostate cancer. J Urol. 1992;147(3 Pt 2):846-51; discussion 51-2.
97. Brawer MK, Chetner MP, Beatie J, Buchner DM, Vessella RL, Lange PH. Screening for prostatic carcinoma with prostate specific antigen. J Urol. 1992;147(3 Pt 2):841-5.
98. Catalona WJ, Richie JP, Ahmann FR, et al. Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men. J Urol. 1994;151(5):1283-90.
99. National Cancer Institute at the National Institutes of Health. Fact Sheets. Available from: http://www.cancer.gov/cancertopics/factsheet/detection/PSA. [Accessed 6 Oct 2014]
100. World Health Organization, WHO. Early detection of cancer. Available from: http://www.who.int/cancer/detection/en/ [Accessed 6 Oct 2014]
101. International Agency for Research on Cancer WHO. Cancer Epidemiology: Principles and Methods. 1999. Available from: http://www.iarc.fr/en/publications/pdfs-online/epi/cancerepi/. [Accessed 6 Oct 2014]
102. Rothman KJ, Greenland S. Modern Epidemiology. Second Edition ed. Philadelphia: Lippincott-Rave Publishers; 1998.
103. Aschengrau A, Seage GRI. Essentials of Epidemiology in Public Health. Second Edition ed. Sudbury, Massachusetts: Jones and Bartlett Publisher; 2008.
104. Wilson JMG, Jungner G. Principles and Practice of Screening for Disease. World Health Organization; 1968.
105. Andermann A, Blancquaert I, Beauchamp S, Dery V. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. Bulletin of the World Health Organization. 2008;86(4):317-9.
106. The Pennsylvania State University. 2014; Available from: https://onlinecourses.science.psu.edu/stat507/node/74. [Accessed 9 Oct 2014]
107. Moyer VA. Screening for Prostate Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2012.
108. Ilic D, Neuberger MM, Djulbegovic M, Dahm P. Screening for prostate cancer. Cochrane Database Syst Rev. 2013;1:CD004720.

109. The National Board of Health and Welfare S. Nationella Screeningprogram.; Available from: http://www.socialstyrelsen.se/riktlinjer/nationellascreeningprogram/prostatacancer-screeningmedpsa. [Accessed 20 Oct 2014]
110. American Urological Association. Guidelines: Detection of Prostate Cancer; Available from: https://www.auanet.org/education/guidelines/prostate-cancer-detection.cfm. [Accessed 9 Nov 2014]
111. Carroll PR, Parsons JK, Andriole G, et al. Prostate cancer early detection, version 1.2014. Journal of the National Comprehensive Cancer Network : JNCCN. 2014;12(9):1211-9.
112. Catalona WJ, D'Amico AV, Fitzgibbons WF, et al. What the U.S. Preventive Services Task Force missed in its prostate cancer screening recommendation. Ann Intern Med. 2012;157(2):137-8.
113. Carlsson S, Vickers AJ, Roobol M, et al. Prostate cancer screening: facts, statistics, and interpretation in response to the US Preventive Services Task Force Review. J Clin Oncol. 2012;30(21):2581-4.
114. Schroder FH, van der Maas P, Beemsterboer P, et al. Evaluation of the digital rectal examination as a screening test for prostate cancer. Rotterdam section of the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst. 1998;90(23):1817-23.
115. Gosselaar C, Roobol MJ, Roemeling S, Schroder FH. The role of the digital rectal examination in subsequent screening visits in the European randomized study of screening for prostate cancer (ERSPC), Rotterdam. Eur Urol. 2008;54(3):581-8.
116. Thompson IM, Ankerst DP, Chi C, et al. Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial. J Natl Cancer Inst. 2006;98(8):529-34.
117. Okotie OT, Roehl KA, Han M, Loeb S, Gashti SN, Catalona WJ. Characteristics of prostate cancer detected by digital rectal examination only. Urology. 2007;70(6):1117-20.
118. Eastham JA, Riedel E, Scardino PT, et al. Variation of serum prostate-specific antigen levels: an evaluation of year-to-year fluctuations. JAMA. 2003;289(20):2695-700.
119. Hugosson J, Aus G, Bergdahl S, et al. Population-based screening for prostate cancer by measuring free and total serum prostate-specific antigen in Sweden. BJU Int. 2003;92 Suppl 2:39-43.
120. Auvinen A, Maattanen L, Finne P, et al. Test sensitivity of prostate-specific antigen in the Finnish randomised prostate cancer screening trial. Int J Cancer. 2004;111(6):940-3.
121. Thompson IM, Ankerst DP, Chi C, et al. Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower. JAMA. 2005;294(1):66-70.

122. Thompson IM, Pauler DK, Goodman PJ, et al. Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. N Engl J Med. 2004;350(22):2239-46.
123. Maattanen L, Hakama M, Tammela TL, et al. Specificity of serum prostate-specific antigen determination in the Finnish prostate cancer screening trial. Br J Cancer. 2007;96(1):56-60.
124. Wolf AM, Wender RC, Etzioni RB, et al. American Cancer Society guideline for the early detection of prostate cancer: update 2010. CA: a cancer journal for clinicians. 2010;60(2):70-98.
125. Benson MC, Whang IS, Pantuck A, et al. Prostate specific antigen density: a means of distinguishing benign prostatic hypertrophy and prostate cancer. J Urol. 1992;147(3 Pt 2):815-6.
126. Catalona WJ, Southwick PC, Slawin KM, et al. Comparison of percent free PSA, PSA density, and age-specific PSA cutoffs for prostate cancer detection and staging. Urology. 2000;56(2):255-60.
127. Catalona WJ, Partin AW, Slawin KM, et al. Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: a prospective multicenter clinical trial. JAMA. 1998;279(19):1542-7.
128. Lee R, Localio AR, Armstrong K, Malkowicz SB, Schwartz JS. A meta-analysis of the performance characteristics of the free prostate-specific antigen test. Urology. 2006;67(4):762-8.
129. Vickers AJ, Savage C, O'Brien MF, Lilja H. Systematic review of pretreatment prostate-specific antigen velocity and doubling time as predictors for prostate cancer. J Clin Oncol. 2009;27(3):398-403.
130. Vickers AJ, Thompson IM, Klein E, Carroll PR, Scardino PT. A commentary on PSA velocity and doubling time for clinical decisions in prostate cancer. Urology. 2014;83(3):592-6.
131. Trabulsi EJ, Sackett D, Gomella LG, Halpern EJ. Enhanced transrectal ultrasound modalities in the diagnosis of prostate cancer. Urology. 2010;76(5):1025-33.
132. Bjurlin MA, Carter HB, Schellhammer P, et al. Optimization of initial prostate biopsy in clinical practice: sampling, labeling and specimen processing. J Urol. 2013;189(6):2039-46.
133. Amercian Urological Association Inc. AUA/Optimal Technique of Prostate Biopsy and Specimen Handling. 2013. http://www.auanet.org/common/pdf/education/clinical-guidance/Prostate-Biopsy-WhitePaper.pdf [Accessed 6 Oct 20149
134. Berglund RK, Masterson TA, Vora KC, Eggener SE, Eastham JA, Guillonneau BD. Pathological upgrading and up staging with immediate repeat biopsy in patients eligible for active surveillance. J Urol. 2008;180(5):1964-7; discussion 7-8.
135. Porten SP, Whitson JM, Cowan JE, et al. Changes in prostate cancer grade on serial biopsy in men undergoing active surveillance. J Clin Oncol. 2011;29(20):2795-800.

136. Motamedinia P, RiChard JL, McKiernan JM, DeCastro GJ, Benson MC. Role of immediate confirmatory prostate biopsy to ensure accurate eligibility for active surveillance. Urology. 2012;80(5):1070-4.
137. Kahl P, Wolf S, Adam A, et al. Saturation biopsy improves preoperative Gleason scoring of prostate cancer. Pathology, research and practice. 2009;205(4):259-64.
138. Barzell WE, Melamed MR, Cathcart P, Moore CM, Ahmed HU, Emberton M. Identifying candidates for active surveillance: an evaluation of the repeat biopsy strategy for men with favorable risk prostate cancer. J Urol. 2012;188(3):762-7.
139. Berglund RK, Jones JS. Optimal prostate biopsy regimen. Current urology reports. 2012;13(6):455-9.
140. Eichler K, Hempel S, Wilby J, Myers L, Bachmann LM, Kleijnen J. Diagnostic value of systematic biopsy methods in the investigation of prostate cancer: a systematic review. J Urol. 2006;175(5):1605-12.
141. Labrie F, Candas B, Cusan L, et al. Screening decreases prostate cancer mortality: 11-year follow-up of the 1988 Quebec prospective randomized controlled trial. Prostate. 2004;59(3):311-8.
142. Labrie F, Candas B, Dupont A, et al. Screening decreases prostate cancer death: first analysis of the 1988 Quebec prospective randomized controlled trial. Prostate. 1999;38(2):83-91.
143. Sandblom G, Varenhorst E, Lofman O, Rosell J, Carlsson P. Clinical consequences of screening for prostate cancer: 15 years follow-up of a randomised controlled trial in Sweden. Eur Urol. 2004;46(6):717-23; discussion 24.
144. Kjellman A, Akre O, Norming U, Tornblom M, Gustafsson O. 15-year followup of a population based prostate cancer screening study. J Urol. 2009;181(4):1615-21; discussion 21.
145. Gohagan JK, Prorok PC, Hayes RB, Kramer BS, Prostate LC, Ovarian Cancer Screening Trial Project T. The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: history, organization, and status. Control Clin Trials. 2000;21(6 Suppl):251S-72S.
146. Andriole GL, Crawford ED, Grubb RL, 3rd, et al. Mortality results from a randomized prostate-cancer screening trial. N Engl J Med. 2009;360(13):1310-9.
147. Andriole GL, Crawford ED, Grubb RL, 3rd, et al. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. J Natl Cancer Inst. 2012;104(2):125-32.
148. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. N Engl J Med. 2009;360(13):1320-8.

149. Schroder FH, Hugosson J, Roobol MJ, et al. Prostate-cancer mortality at 11 years of follow-up. N Engl J Med. 2012;366(11):981-90.
150. Pinsky PF, Blacka A, Kramer BS, Miller A, Prorok PC, Berg C. Assessing contamination and compliance in the prostate component of the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. Clin Trials. 2010;7(4):303-11.
151. Grubb RL, 3rd, Pinsky PF, Greenlee RT, et al. Prostate cancer screening in the Prostate, Lung, Colorectal and Ovarian cancer screening trial: update on findings from the initial four rounds of screening in a randomized trial. BJU Int. 2008;102(11):1524-30.
152. Schroder FH, Roobol MJ. ERSPC and PLCO prostate cancer screening studies: what are the differences? Eur Urol. 2010;58(1):46-52.
153. Barry MJ. Screening for prostate cancer--the controversy that refuses to die. N Engl J Med. 2009;360(13):1351-4.
154. Epstein MM, Edgren G, Rider JR, Mucci LA, Adami HO. Temporal trends in cause of death among Swedish and US men with prostate cancer. J Natl Cancer Inst. 2012;104(17):1335-42.
155. Draisma G, Etzioni R, Tsodikov A, et al. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. J Natl Cancer Inst. 2009;101(6):374-83.
156. Pearson JD, Luderer AA, Metter EJ, et al. Longitudinal analysis of serial measurements of free and total PSA among men with and without prostatic cancer. Urology. 1996;48(6A Suppl):4-9.
157. Tornblom M, Eriksson H, Franzen S, et al. Lead time associated with screening for prostate cancer. Int J Cancer. 2004;108(1):122-9.
158. Savage CJ, Lilja H, Cronin AM, Ulmert D, Vickers AJ. Empirical estimates of the lead time distribution for prostate cancer based on two independent representative cohorts of men not subject to prostate-specific antigen screening. Cancer Epidemiol Biomarkers Prev. 2010;19(5):1201-7.
159. Auvinen A, Maattanen L, Stenman UH, et al. Lead-time in prostate cancer screening (Finland). Cancer Causes Control. 2002;13(3):279-85.
160. Pashayan N, Duffy SW, Pharoah P, et al. Mean sojourn time, overdiagnosis, and reduction in advanced stage prostate cancer due to screening with PSA: implications of sojourn time on screening. Br J Cancer. 2009;100(7):1198-204.
161. Draisma G, van Rosmalen J. A note on the catch-up time method for estimating lead or sojourn time in prostate cancer screening. Stat Med. 2013;32(19):3332-41.
162. Finne P, Fallah M, Hakama M, et al. Lead-time in the European Randomised Study of Screening for Prostate Cancer. Eur J Cancer. 2010;46(17):3102-8.

163. Telesca D, Etzioni R, Gulati R. Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends. Biometrics. 2008;64(1):10-9.
164. Etzioni R, Penson DF, Legler JM, et al. Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends. J Natl Cancer Inst. 2002;94(13):981-90.
165. Wu GH, Auvinen A, Maattanen L, et al. Number of screens for overdetection as an indicator of absolute risk of overdiagnosis in prostate cancer screening. Int J Cancer. 2012;131(6):1367-75.
166. Zappa M, Ciatto S, Bonardi R, Mazzotta A. Overdiagnosis of prostate carcinoma by screening: an estimate based on the results of the Florence Screening Pilot Study. Ann Oncol. 1998;9(12):1297-300.
167. Pashayan N, Powles J, Brown C, Duffy SW. Excess cases of prostate cancer and estimated overdiagnosis associated with PSA testing in East Anglia. Br J Cancer. 2006;95(3):401-5.
168. McGregor M, Hanley JA, Boivin JF, McLean RG. Screening for prostate cancer: estimating the magnitude of overdetection. CMAJ. 1998;159(11):1368-72.
169. Ciatto S, Gervasi G, Bonardi R, et al. Determining overdiagnosis by screening with DRE/TRUS or PSA (Florence pilot studies, 1991-1994). Eur J Cancer. 2005;41(3):411-5.
170. Heijnsdijk EA, Wever EM, Auvinen A, et al. Quality-of-life effects of prostate-specific antigen screening. N Engl J Med. 2012;367(7):595-605.
171. Gulati R, Gore JL, Etzioni R. Comparative effectiveness of alternative prostate-specific antigen--based prostate cancer screening strategies: model estimates of potential benefits and harms. Ann Intern Med. 2013;158(3):145-53.
172. Gulati R, Inoue LY, Gore JL, Katcher J, Etzioni R. Individualized estimates of overdiagnosis in screen-detected prostate cancer. J Natl Cancer Inst. 2014;106(2):djt367.
173. de Carvalho TM, Heijnsdijk EA, de Koning HJ. Screening for prostate cancer in the US? reduce the harms and keep the benefit. Int J Cancer. 2014.
174. 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(11):831-8.
175. de Gelder R, Heijnsdijk EA, van Ravesteyn NT, Fracheboud J, Draisma G, de Koning HJ. Interpreting overdiagnosis estimates in population-based mammography screening. Epidemiologic reviews. 2011;33(1):111-21.
176. Welch HG, Albertsen PC. Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986-2005. J Natl Cancer Inst. 2009;101(19):1325-9.

177. Gulati R, Mariotto AB, Chen S, Gore JL, Etzioni R. Long-term projections of the harm-benefit trade-off in prostate cancer screening are more favorable than previous short-term estimates. Journal of clinical epidemiology. 2011;64(12):1412-7.
178. Loeb S, Vonesh EF, Metter EJ, Carter HB, Gann PH, Catalona WJ. What is the true number needed to screen and treat to save a life with prostate-specific antigen testing? J Clin Oncol. 2011;29(4):464-7.
179. Kilpelainen TP, Tammela TL, Malila N, et al. Prostate cancer mortality in the Finnish randomized screening trial. J Natl Cancer Inst. 2013;105(10):719-25.
180. Roobol MJ, Kranse R, Bangma CH, et al. Screening for prostate cancer: results of the rotterdam section of the European randomized study of screening for prostate cancer. Eur Urol. 2013;64(4):530-9.
181. Ploussard G, Epstein JI, Montironi R, et al. The contemporary concept of significant versus insignificant prostate cancer. Eur Urol. 2011;60(2):291-303.
182. Stamey TA, Freiha FS, McNeal JE, Redwine EA, Whittemore AS, Schmid HP. Localized prostate cancer. Relationship of tumor volume to clinical significance for treatment of prostate cancer. Cancer. 1993;71(3 Suppl):933-8.
183. Epstein JI, Walsh PC, Carmichael M, Brendler CB. Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer. JAMA. 1994;271(5):368-74.
184. Andreoiu M, Cheng L. Multifocal prostate cancer: biologic, prognostic, and therapeutic implications. Human pathology. 2010;41(6):781-93.
185. Wolters T, Roobol MJ, van Leeuwen PJ, et al. A critical analysis of the tumor volume threshold for clinically insignificant prostate cancer using a data set of a randomized screening trial. J Urol. 2011;185(1):121-5.
186. Bangma CH, Roobol MJ. Defining and predicting indolent and low risk prostate cancer. Crit Rev Oncol Hematol. 2011.
187. Bastian PJ, Mangold LA, Epstein JI, Partin AW. Characteristics of insignificant clinical T1c prostate tumors. A contemporary analysis. Cancer. 2004;101(9):2001-5.
188. Steyerberg EW, Roobol MJ, Kattan MW, van der Kwast TH, de Koning HJ, Schroder FH. Prediction of indolent prostate cancer: validation and updating of a prognostic nomogram. J Urol. 2007;177(1):107-12; discussion 12.
189. Kattan MW, Eastham JA, Wheeler TM, et al. Counseling men with prostate cancer: a nomogram for predicting the presence of small, moderately differentiated, confined tumors. J Urol. 2003;170(5):1792-7.
190. Nakanishi H, Wang X, Ochiai A, et al. A nomogram for predicting low-volume/low-grade prostate cancer: a tool in selecting patients for active surveillance. Cancer. 2007;110(11):2441-7.

191. Bangma CH, Roobol MJ, Steyerberg EW. Predictive models in diagnosing indolent cancer. Cancer. 2009;115(13 Suppl):3100-6.
192. prostatacancerregistret. N. Nationell kvalitetsrapport för diagnosår 2012. 2012.
193. Damiano R, Di Lorenzo G, Cantiello F, et al. Clinicopathologic features of prostate adenocarcinoma incidentally discovered at the time of radical cystectomy: an evidence-based analysis. Eur Urol. 2007;52(3):648-57.
194. Postma R, Schroder FH, van Leenders GJ, et al. Cancer detection and cancer characteristics in the European Randomized Study of Screening for Prostate Cancer (ERSPC)--Section Rotterdam. A comparison of two rounds of screening. Eur Urol. 2007;52(1):89-97.
195. Statistics Sweden Statistiska Centralbyrån. Befolkningsstatistik. Avalable from: http://www.scb.se/sv_/Hitta-statistik/Statistik-efter-amne/Befolkning/Befolkningens-sammansattning/Befolkningsstatistik/25788/25795/Helarsstatistik---Riket/25830/. [Accessed 9 Nov 2014]
196. Social Security. Official Social Security Website. Available from: www.ssa.gov/oact/STATS. [Accessed 9 Nov 2014]
197. Walter LC, Covinsky KE. Cancer screening in elderly patients: a framework for individualized decision making. JAMA. 2001;285(21):2750-6.
198. Vickers AJ, Sjoberg DD, Ulmert D, et al. Empirical estimates of prostate cancer overdiagnosis by age and prostate-specific antigen. BMC medicine. 2014;12:26.
199. Loeb S, Vellekoop A, Ahmed HU, et al. Systematic Review of Complications of Prostate Biopsy. Eur Urol. 2013.
200. Rosario DJ, Lane JA, Metcalfe C, et al. Short term outcomes of prostate biopsy in men tested for cancer by prostate specific antigen: prospective evaluation within ProtecT study. BMJ. 2012;344:d7894.
201. Mkinen T, Auvinen A, Hakama M, Stenman UH, Tammela TL. Acceptability and complications of prostate biopsy in population-based PSA screening versus routine clinical practice: a prospective, controlled study. Urology. 2002;60(5):846-50.
202. Nam RK, Saskin R, Lee Y, et al. Increasing hospital admission rates for urological complications after transrectal ultrasound guided prostate biopsy. J Urol. 2013;189(1 Suppl):S12-7; discussion S7-8.
203. Lundstrom KJ, Drevin L, Carlsson S, et al. Nationwide population-based study of infections after transrectal ultrasound guided prostate biopsy. J Urol. 2014.
204. Loeb S, van den Heuvel S, Zhu X, Bangma CH, Schroder FH, Roobol MJ. Infectious complications and hospital admissions after prostate biopsy in a European randomized trial. Eur Urol. 2012;61(6):1110-4.

205. Loeb S, Carter HB, Berndt SI, Ricker W, Schaeffer EM. Complications after prostate biopsy: data from SEER-Medicare. J Urol. 2011;186(5):1830-4.
206. Pinsky PF, Parnes HL, Andriole G. Mortality and Complications Following Prostate Biopsy in the PLCO Cancer Screening Trial. BJU Int. 2013.
207. Carlsson SV, Holmberg E, Moss SM, et al. No excess mortality after prostate biopsy: results from the European Randomized Study of Screening for Prostate Cancer. BJU Int. 2011;107(12):1912-7.
208. Hewitson P, Austoker J. Part 2: Patient information, informed decision-making and the psycho-social impact of prostate-specific antigen testing. BJU Int. 2005;95 Suppl 3:16-32.
209. Carlsson S, Aus G, Wessman C, Hugosson J. Anxiety associated with prostate cancer screening with special reference to men with a positive screening test (elevated PSA) - Results from a prospective, population-based, randomised study. Eur J Cancer. 2007;43(14):2109-16.
210. Brindle LA, Oliver SE, Dedman D, et al. Measuring the psychosocial impact of population-based prostate-specific antigen testing for prostate cancer in the UK. BJU Int. 2006;98(4):777-82.
211. Macefield RC, Lane JA, Metcalfe C, et al. Do the risk factors of age, family history of prostate cancer or a higher prostate specific antigen level raise anxiety at prostate biopsy? Eur J Cancer. 2009;45(14):2569-73.
212. Wade J, Rosario DJ, Macefield RC, et al. Psychological impact of prostate biopsy: physical symptoms, anxiety, and depression. J Clin Oncol. 2013;31(33):4235-41.
213. Carlsson S, Sandin F, Fall K, et al. Risk of suicide in men with low-risk prostate cancer. Eur J Cancer. 2013;49(7):1588-99.
214. Fang F, Keating NL, Mucci LA, et al. Immediate risk of suicide and cardiovascular death after a prostate cancer diagnosis: cohort study in the United States. J Natl Cancer Inst. 2010;102(5):307-14.
215. Bill-Axelson A, Garmo H, Nyberg U, et al. Psychiatric treatment in men with prostate cancer--results from a Nation-wide, population-based cohort study from PCBaSe Sweden. Eur J Cancer. 2011;47(14):2195-201.
216. Fall K, Fang F, Mucci LA, et al. Immediate risk for cardiovascular events and suicide following a prostate cancer diagnosis: prospective cohort study. PLoS medicine. 2009;6(12):e1000197.
217. McNaughton-Collins M, Fowler FJ, Jr., Caubet JF, et al. Psychological effects of a suspicious prostate cancer screening test followed by a benign biopsy result. The American journal of medicine. 2004;117(10):719-25.
218. Fowler FJ, Jr., Barry MJ, Walker-Corkery B, et al. The impact of a suspicious prostate biopsy on patients' psychological, socio-behavioral, and medical care outcomes. Journal of general internal medicine. 2006;21(7):715-21.

219. Heijnsdijk EA, der Kinderen A, Wever EM, Draisma G, Roobol MJ, de Koning HJ. Overdetection, overtreatment and costs in prostate-specific antigen screening for prostate cancer. Br J Cancer. 2009;101(11):1833-8.
220. Punnen S, Cowan JE, Chan JM, Carroll PR, Cooperberg MR. Long-term Health-related Quality of Life After Primary Treatment for Localized Prostate Cancer: Results from the CaPSURE Registry. Eur Urol. 2014.
221. Potosky AL, Legler J, Albertsen PC, et al. Health outcomes after prostatectomy or radiotherapy for prostate cancer: results from the Prostate Cancer Outcomes Study. J Natl Cancer Inst. 2000;92(19):1582-92.
222. Resnick MJ, Koyama T, Fan KH, et al. Long-term functional outcomes after treatment for localized prostate cancer. N Engl J Med. 2013;368(5):436-45.
223. Alivizatos G, Skolarikos A. Incontinence and erectile dysfunction following radical prostatectomy: a review. ScientificWorldJournal. 2005;5:747-58.
224. Walsh PC, Lepor H, Eggleston JC. Radical prostatectomy with preservation of sexual function: anatomical and pathological considerations. Prostate. 1983;4(5):473-85.
225. Johansson E, Steineck G, Holmberg L, et al. Long-term quality-of-life outcomes after radical prostatectomy or watchful waiting: the Scandinavian Prostate Cancer Group-4 randomised trial. Lancet Oncol. 2011.
226. Carlsson S, Aus G, Bergdahl S, et al. The excess burden of side-effects from treatment in men allocated to screening for prostate cancer. The Goteborg randomised population-based prostate cancer screening trial. Eur J Cancer. 2010.
227. Carlsson S, Adolfsson J, Bratt O, et al. Nationwide population-based study on 30-day mortality after radical prostatectomy in Sweden. Scand J Urol Nephrol. 2009;43(5):350-6.
228. Ataman F, Zurlo A, Artignan X, et al. Late toxicity following conventional radiotherapy for prostate cancer: analysis of the EORTC trial 22863. Eur J Cancer. 2004;40(11):1674-81.
229. Sanda MG, Dunn RL, Michalski J, et al. Quality of life and satisfaction with outcome among prostate-cancer survivors. N Engl J Med. 2008;358(12):1250-61.
230. Schaake W, Wiegman EM, de Groot M, et al. The impact of gastrointestinal and genitourinary toxicity on health related quality of life among irradiated prostate cancer patients. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 2014;110(2):284-90.
231. Davis KM, Kelly SP, Luta G, Tomko C, Miller AB, Taylor KL. The association of long-term treatment-related side effects with cancer-specific and general quality of life among prostate cancer survivors. Urology. 2014;84(2):300-6.

232. Wright JM, Musini VM. First-line drugs for hypertension. Cochrane Database Syst Rev. 2009(3):CD001841.
233. Gram IT, Lund E, Slenker SE. Quality of life following a false positive mammogram. Br J Cancer. 1990;62(6):1018-22.
234. Gotzsche PC, Nielsen M. Screening for breast cancer with mammography. Cochrane Database Syst Rev. 2011(1):CD001877.
235. Welch HG, Black WC. Using autopsy series to estimate the disease "reservoir" for ductal carcinoma in situ of the breast: how much more breast cancer can we find? Ann Intern Med. 1997;127(11):1023-8.
236. Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L. Screening for breast cancer: an update for the U.S. Preventive Services Task Force. Ann Intern Med. 2009;151(10):727-37, W237-42.
237. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2009;151(10):716-26, W-236.
238. The benefits and harms of breast cancer screening: an independent review. Lancet. 2012;380(9855):1778-86.
239. The National Board of Health and Welfare. Screening för bröstcancer. Rekommendation och bedömningsunderlag. Remissversion 2013. Available from: http://www.socialstyrelsen.se/SiteCollectionDocuments/nr-screening-brostcancer-rekommendation.pdf [Accessed 9 Nov 2014]
240. Det Nordiske Cochrane center, 2012; Available from: http://www.cochrane.dk/screening/mammografi-svensk.pdf. [Accessed 9 nov 2014]
241. Holme O, Bretthauer M, Fretheim A, Odgaard-Jensen J, Hoff G. Flexible sigmoidoscopy versus faecal occult blood testing for colorectal cancer screening in asymptomatic individuals. Cochrane Database Syst Rev. 2013;9:CD009259.
242. Shaukat A, Mongin SJ, Geisser MS, et al. Long-term mortality after screening for colorectal cancer. N Engl J Med. 2013;369(12):1106-14.
243. Scholefield JH, Moss SM, Mangham CM, Whynes DK, Hardcastle JD. Nottingham trial of faecal occult blood testing for colorectal cancer: a 20-year follow-up. Gut. 2012;61(7):1036-40.
244. Kronborg O, Jorgensen OD, Fenger C, Rasmussen M. Randomized study of biennial screening with a faecal occult blood test: results after nine screening rounds. Scandinavian journal of gastroenterology. 2004;39(9):846-51.
245. Kewenter J, Brevinge H, Engaras B, Haglind E, Ahren C. Results of screening, rescreening, and follow-up in a prospective randomized study for detection of colorectal cancer by fecal occult blood testing. Results for 68,308 subjects. Scandinavian journal of gastroenterology. 1994;29(5):468-73.
246. Hewitson P, Glasziou P, Watson E, Towler B, Irwig L. Cochrane systematic review of colorectal cancer screening using the fecal

occult blood test (hemoccult): an update. The American journal of gastroenterology. 2008;103(6):1541-9.
247. Ekelund G, Manjer J, Zackrisson S. Population-based screening for colorectal cancer with faecal occult blood test--do we really have enough evidence? International journal of colorectal disease. 2010;25(11):1269-75.
248. Hoff G, Grotmol T, Skovlund E, Bretthauer M. Risk of colorectal cancer seven years after flexible sigmoidoscopy screening: randomised controlled trial. BMJ. 2009;338:b1846.
249. Schoen RE, Pinsky PF, Weissfeld JL, et al. Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. N Engl J Med. 2012;366(25):2345-57.
250. Atkin WS, Edwards R, Kralj-Hans I, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. Lancet. 2010;375(9726):1624-33.
251. Thiis-Evensen E, Hoff GS, Sauar J, Langmark F, Majak BM, Vatn MH. Population-based surveillance by colonoscopy: effect on the incidence of colorectal cancer. Telemark Polyp Study I. Scandinavian journal of gastroenterology. 1999;34(4):414-20.
252. Segnan N, Armaroli P, Bonelli L, et al. Once-only sigmoidoscopy in colorectal cancer screening: follow-up findings of the Italian Randomized Controlled Trial--SCORE. J Natl Cancer Inst. 2011;103(17):1310-22.
253. The National Board of Health and Welfare S. Tjock- och ändtarmscancer, screening med test av blod i avföringen. Available from http://www.socialstyrelsen.se/SiteCollectionDocuments/screening-tjockandtarmscancer-rekommendation.pdf [Accessed Nov 2014]
254. Fidler HK, Boyes DA, Worth AJ. Cervical cancer detection in British Columbia. A progress report. The Journal of obstetrics and gynaecology of the British Commonwealth. 1968;75(4):392-404.
255. Hakama M, Rasanen-Virtanen U. Effect of a mass screening program on the risk of cervical cancer. American journal of epidemiology. 1976;103(5):512-7.
256. Laara E, Day NE, Hakama M. Trends in mortality from cervical cancer in the Nordic countries: association with organised screening programmes. Lancet. 1987;1(8544):1247-9.
257. Nieminen P, Kallio M, Anttila A, Hakama M. Organised vs. spontaneous Pap-smear screening for cervical cancer: A case-control study. Int J Cancer. 1999;83(1):55-8.
258. Gates TJ. Screening for cancer: evaluating the evidence. American family physician. 2001;63(3):513-22.
259. Sharp L, Cotton S, Cochran C, et al. After-effects reported by women following colposcopy, cervical biopsies and LLETZ: results from the TOMBOLA trial. BJOG : an international journal of obstetrics and gynaecology. 2009;116(11):1506-14.

260. Kyrgiou M, Koliopoulos G, Martin-Hirsch P, Arbyn M, Prendiville W, Paraskevaidis E. Obstetric outcomes after conservative treatment for intraepithelial or early invasive cervical lesions: systematic review and meta-analysis. Lancet. 2006;367(9509):489-98.
261. Moyer VA. Screening for cervical cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2012;156(12):880-91, W312.
262. Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. European journal of epidemiology. 2009;24(11):659-67.
263. The National Board of Health and Welfare S. Dödsorsaksregistret. Available from: http://www.socialstyrelsen.se/register/dodsorsaksregistret. [Accessed 6 Oct 2014]
264. The National Board of Health and Welfare S. Dödsorsaksstatistik – Historik, produktionsmetoder och tillförlitlighet. 2010; Available from: http://www.socialstyrelsen.se/publikationer2010/2010-4-33. [Accessed 6 Oct 2014]
265. The National Board of Health and Welfare S. Dödsorsaker 2013. 2014; Available from: http://www.socialstyrelsen.se/publikationer2014/2014-8-5. [Accessed 6 Oct 2014]
266. The National Board of Health and Welfare S. Cancerregistret. [Sept 29 2014]; Available from: http://www.socialstyrelsen.se/register/halsodataregister/cancerregistret. [Accessed 6 Oct 2014]
267. Barlow L, Westergren K, Holmberg L, Talback M. The completeness of the Swedish Cancer Register: a sample survey for year 1998. Acta Oncol. 2009;48(1):27-33.
268. De Koning HJ, Blom J, Merkelbach JW, et al. Determining the cause of death in randomized screening trial(s) for prostate cancer. BJU Int. 2003;92 Suppl 2:71-8.
269. Jonsson H, Holmstrom B, Duffy SW, Stattin P. Uptake of prostate-specific antigen testing for early prostate cancer detection in Sweden. Int J Cancer. 2011;129(8):1881-8.
270. Nordstrom T, Aly M, Clements MS, Weibull CE, Adolfsson J, Gronberg H. Prostate-specific antigen (PSA) testing is prevalent and increasing in Stockholm County, Sweden, Despite no recommendations for PSA screening: results from a population-based study, 2003-2011. Eur Urol. 2013;63(3):419-25.
271. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Family medicine. 2005;37(5):360-3.
272. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med. 1998;17(8):857-72.

273. Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc. 1927;22:209-12.
274. Fleiss FL. Statistical methods for rates and proportions. 2nd ed. New York: Wiley; 1981.
275. Altman DG. Practical Statistics for Medical Research. 1st ed. London: Chapman and Hall; 1991.
276. Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). BMJ. 1998;317(7172):1572.
277. Vittinghoff E, Glidden DV, Shiboski S, McCulloch CE. Regression methods in biostatistics : linear, logistic, survival, and repeated measures models New York: Springer Science+Business Media, Inc.; 2005.
278. Greenwood M. The Errors of Sampling of the Survivorship Table Reports on Public Health and Medical Subjects. 1926;33.
279. Wolkewitz M, Cooper BS, Bonten MJ, Barnett AG, Schumacher M. Interpreting and comparing risks in the presence of competing events. BMJ. 2014;349:g5060.
280. Kim HT. Cumulative incidence in competing risks data and competing risks regression analysis. Clinical cancer research : an official journal of the American Association for Cancer Research. 2007;13(2 Pt 1):559-65.
281. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statisical Association. 1999;94(446):496-509.
282. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: a statistical methodology. Natl Cancer Inst Monogr. 1961;6:101-21.
283. Dickman PW, Coviello E, Hills M. Estimating and modelling relative survival. The Stata Journal. 2013:1-25.
284. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. BMJ. 1995;310(6977):452-4.
285. Rembold CM. Number needed to screen: development of a statistic for disease screening. BMJ. 1998;317(7154):307-12.
286. Suissa S. Calculation of number needed to treat. N Engl J Med. 2009;361(4):424-5.
287. Hildebrandt M, Vervolgyi E, Bender R. Calculation of NNTs in RCTs with time-to-event outcomes: a literature review. BMC medical research methodology. 2009;9:21.
288. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. BMJ. 1999;319(7223):1492-5.
289. Chamberlain J, Coleman D, Ellman R, Moss S. Verification of the cause of death in the trial of early detection of breast cancer. UK Trial of Early Detection of Breast Cancer Group. Trial Co-ordinating Centre. Br J Cancer. 1991;64(6):1151-6.
290. Doria-Rose VP, Marcus PM, Miller AB, et al. Does the source of death information affect cancer screening efficacy results? A study of the

use of mortality review versus death certificates in four randomized trials. Clin Trials. 2010;7(1):69-77.
291. Miller AB, Yurgalevitch S, Weissfeld JL. Death review process in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. Control Clin Trials. 2000;21(6 Suppl):400S-6S.
292. Nystrom L, Larsson LG, Rutqvist LE, et al. Determination of cause of death among breast cancer cases in the Swedish randomized mammography screening trials. A comparison between official statistics and validation by an endpoint committee. Acta Oncol. 1995;34(2):145-52.
293. Robinson MH, Rodrigues VC, Hardcastle JD, Chamberlain JO, Mangham CM, Moss SM. Faecal occult blood screening for colorectal cancer at Nottingham: details of the verification process. J Med Screen. 2000;7(2):97-8.
294. Otto SJ, van Leeuwen PJ, Hoekstra JW, et al. Blinded and uniform causes of death verification in cancer screening: a major influence on the outcome of a prostate cancer screening trial? Eur J Cancer. 2010;46(17):3061-7.
295. Makinen T, Karhunen P, Aro J, Lahtela J, Maattanen L, Auvinen A. Assessment of causes of death in a prostate cancer screening trial. Int J Cancer. 2008;122(2):413-7.
296. Fall K, Stromberg F, Rosell J, Andren O, Varenhorst E. Reliability of death certificates in prostate cancer patients. Scand J Urol Nephrol. 2008;42(4):352-7.
297. Black WC, Haggstrom DA, Welch HG. All-cause mortality in randomized trials of cancer screening. J Natl Cancer Inst. 2002;94(3):167-73.
298. Feuer EJ, Merrill RM, Hankey BF. Cancer surveillance series: interpreting trends in prostate cancer--part II: Cause of death misclassification and the recent rise and fall in prostate cancer mortality. J Natl Cancer Inst. 1999;91(12):1025-32.
299. Hoffman RM, Stone SN, Hunt WC, Key CR, Gilliland FD. Effects of misattribution in assigning cause of death on prostate cancer mortality rates. Annals of epidemiology. 2003;13(6):450-4.
300. Albertsen PC, Walters S, Hanley JA. A comparison of cause of death determination in men previously diagnosed with prostate cancer who died in 1985 or 1995. J Urol. 2000;163(2):519-23.
301. The National Board of Health and Welfare S. Nationella riktlinjer för bröst-, prostata-, tjocktarms- och ändtarmscancervård 2014. 2014; Available from: http://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/19383/2014-4-2.pdf. [Accessed 6 Nov 2014]
302. Kim SP, Gross CP, Nguyen PL, et al. Perceptions of Active Surveillance and Treatment Recommendations for Low-risk Prostate Cancer: Results from a National Survey of Radiation Oncologists and Urologists. Medical care. 2014;52(7):579-85.

303. Womble PR, Montie JE, Ye Z, et al. Contemporary Use of Initial Active Surveillance Among Men in Michigan with Low-risk Prostate Cancer. Eur Urol. 2014.
304. Iremashvili V, Pelaez L, Manoharan M, Jorda M, Rosenberg DL, Soloway MS. Pathologic prostate cancer characteristics in patients eligible for active surveillance: a head-to-head comparison of contemporary protocols. Eur Urol. 2012;62(3):462-8.
305. Cooperberg MR, Carroll PR, Klotz L. Active surveillance for prostate cancer: progress and promise. J Clin Oncol. 2011;29(27):3669-76.
306. Klotz L, Zhang L, Lam A, Nam R, Mamedov A, Loblaw A. Clinical results of long-term follow-up of a large, active surveillance cohort with localized prostate cancer. J Clin Oncol. 2010;28(1):126-31.
307. Cooperberg MR, Cowan JE, Hilton JF, et al. Outcomes of active surveillance for men with intermediate-risk prostate cancer. J Clin Oncol. 2011;29(2):228-34.
308. Bul M, van den Bergh RC, Zhu X, et al. Outcomes of initially expectantly managed patients with low or intermediate risk screen-detected localized prostate cancer. BJU Int. 2012;110(11):1672-7.
309. Dall'Era MA, Konety BR, Cowan JE, et al. Active surveillance for the management of prostate cancer in a contemporary cohort. Cancer. 2008;112(12):2664-70.
310. Dall'Era MA, Cowan JE, Simko J, et al. Surgical management after active surveillance for low-risk prostate cancer: pathological outcomes compared with men undergoing immediate treatment. BJU Int. 2011;107(8):1232-7.
311. Welty CJ, Cowan JE, Nguyen H, et al. Extended Follow-Up and Risk Factors for Disease Reclassification from a Large Active Surveillance Cohort for Localized Prostate Cancer. J Urol. 2014.
312. van As NJ, Norman AR, Thomas K, et al. Predicting the probability of deferred radical treatment for localised prostate cancer managed by active surveillance. Eur Urol. 2008;54(6):1297-305.
313. Selvadurai ED, Singhera M, Thomas K, et al. Medium-term outcomes of active surveillance for localised prostate cancer. Eur Urol. 2013;64(6):981-7.
314. Soloway MS, Soloway CT, Eldefrawy A, Acosta K, Kava B, Manoharan M. Careful selection and close monitoring of low-risk prostate cancer patients on active surveillance minimizes the need for treatment. Eur Urol. 2010;58(6):831-5.
315. Tosoian JJ, Trock BJ, Landis P, et al. Active surveillance program for prostate cancer: an update of the Johns Hopkins experience. J Clin Oncol. 2011;29(16):2185-90.
316. Adamy A, Yee DS, Matsushita K, et al. Role of prostate specific antigen and immediate confirmatory biopsy in predicting progression during active surveillance for low risk prostate cancer. J Urol. 2011;185(2):477-82.

317. Bul M, Zhu X, Valdagni R, et al. Active surveillance for low-risk prostate cancer worldwide: the PRIAS study. Eur Urol. 2013;63(4):597-603.
318. Klotz L. Active surveillance: the Canadian experience. Curr Opin Urol. 2012;22(3):222-30.
319. Duffield AS, Lee TK, Miyamoto H, Carter HB, Epstein JI. Radical prostatectomy findings in patients in whom active surveillance of prostate cancer fails. J Urol. 2009;182(5):2274-8.
320. Ross AE, Loeb S, Landis P, et al. Prostate-specific antigen kinetics during follow-up are an unreliable trigger for intervention in a prostate cancer surveillance program. J Clin Oncol. 2010;28(17):2810-6.
321. Whitson JM, Porten SP, Hilton JF, et al. The relationship between prostate specific antigen change and biopsy progression in patients on active surveillance for prostate cancer. J Urol. 2011;185(5):1656-60.
322. Mitsuzuka K, Narita S, Koie T, et al. Pathological and biochemical outcomes after radical prostatectomy in men with low-risk prostate cancer meeting the Prostate Cancer International: Active Surveillance criteria. BJU Int. 2013;111(6):914-20.
323. King AC, Livermore A, Laurila TA, Huang W, Jarrard DF. Impact of immediate TRUS rebiopsy in a patient cohort considering active surveillance for favorable risk prostate cancer. Urologic oncology. 2013;31(6):739-43.
324. Ng MK, Van As N, Thomas K, et al. Prostate-specific antigen (PSA) kinetics in untreated, localized prostate cancer: PSA velocity vs PSA doubling time. BJU Int. 2009;103(7):872-6.
325. Al Otaibi M, Ross P, Fahmy N, et al. Role of repeated biopsy of the prostate in predicting disease progression in patients with prostate cancer on active surveillance. Cancer. 2008;113(2):286-92.
326. Venkitaraman R, Norman A, Woode-Amissah R, et al. Predictors of histological disease progression in untreated, localized prostate cancer. J Urol. 2007;178(3 Pt 1):833-7.
327. San Francisco IF, Werner L, Regan MM, Garnick MB, Bubley G, DeWolf WC. Risk stratification and validation of prostate specific antigen density as independent predictor of progression in men with low risk prostate cancer during active surveillance. J Urol. 2011;185(2):471-6.
328. Bul M, van den Bergh RC, Rannikko A, et al. Predictors of unfavourable repeat biopsy results in men participating in a prospective active surveillance program. Eur Urol. 2012;61(2):370-7.
329. Loeb S, Carter HB, Berndt SI, Ricker W, Schaeffer EM. Is repeat prostate biopsy associated with a greater risk of hospitalization? Data from SEER-Medicare. J Urol. 2013;189(3):867-70.
330. Ehdaie B, Vertosick E, Spaliviero M, et al. The Impact of Repeat Biopsies on Infectious Complications in Men with Prostate Cancer on Active Surveillance. J Urol. 2013.

331. Braun K, Ahallal Y, Sjoberg DD, et al. Effect of Repeated Prostate Biopsies on Erectile Function in Men under Active Surveillance for Prostate Cancer. J Urol. 2013.
332. Fujita K, Landis P, McNeil BK, Pavlovich CP. Serial prostate biopsies are associated with an increased risk of erectile dysfunction in men with prostate cancer on active surveillance. J Urol. 2009;182(6):2664-9.
333. Penson DF. Factors influencing patients' acceptance and adherence to active surveillance. Journal of the National Cancer Institute Monographs. 2012;2012(45):207-12.
334. Miocinovic R, Jones JS, Pujara AC, Klein EA, Stephenson AJ. Acceptance and durability of surveillance as a management choice in men with screen-detected, low-risk prostate cancer: improved outcomes with stringent enrollment criteria. Urology. 2011;77(4):980-4.
335. Vasarainen H, Lokman U, Ruutu M, Taari K, Rannikko A. Prostate cancer active surveillance and health-related quality of life: results of the Finnish arm of the prospective trial. BJU Int. 2011.
336. Venderbos LD, van den Bergh RC, Roobol MJ, et al. A longitudinal study on the impact of active surveillance for prostate cancer on anxiety and distress levels. Psycho-oncology. 2014.
337. Loeb S, Folkvaljon Y, Makarov DV, Bratt O, Bill-Axelson A, Stattin P. Five-year Nationwide Follow-up Study of Active Surveillance for Prostate Cancer. Eur Urol. 2014.
338. van den Bergh RC, Roemeling S, Roobol MJ, et al. Outcomes of men with screen-detected prostate cancer eligible for active surveillance who were managed expectantly. Eur Urol. 2009;55(1):1-8.
339. Bellardita L, Rancati T, Alvisi MF, et al. Predictors of Health-related Quality of Life and Adjustment to Prostate Cancer During Active Surveillance. Eur Urol. 2013.
340. Berger ZD, Yeh JC, Carter HB, Pollack CE. Characteristics and Experiences of Patients with Localized Prostate Cancer Who Left an Active Surveillance Program. The patient. 2014.
341. van den Bergh RC, de Blok W, van Muilekom E, Tillier C, Venderbos LD, van der Poel HG. Impact on quality of life of radical prostatectomy after initial active surveillance: more to lose? Scandinavian journal of urology. 2014;48(4):367-73.
342. Sussman R, Staff I, Tortora J, et al. Impact of active surveillance on pathology and nerve sparing status. The Canadian journal of urology. 2014;21(3):7299-304.
343. Bul M, Zhu X, Rannikko A, et al. Radical Prostatectomy for Low-Risk Prostate Cancer Following Initial Active Surveillance: Results From a Prospective Observational Study. Eur Urol. 2012.
344. Khatami A, Aus G, Damber JE, Lilja H, Lodding P, Hugosson J. PSA doubling time predicts the outcome after active surveillance in screening-detected prostate cancer: results from the European randomized

study of screening for prostate cancer, Sweden section. Int J Cancer. 2007;120(1):170-4.
345. Hong SK, Sternberg IA, Keren Paz GE, et al. Definitive Pathology at Radical Prostatectomy Is Commonly Favorable in Men Following Initial Active Surveillance. Eur Urol. 2013.
346. van den Bergh RC, Steyerberg EW, Khatami A, et al. Is delayed radical prostatectomy in men with low-risk screen-detected prostate cancer associated with a higher risk of unfavorable outcomes? Cancer. 2010;116(5):1281-90.
347. Holmstrom B, Holmberg E, Egevad L, et al. Outcome of primary versus deferred radical prostatectomy in the National Prostate Cancer Register of Sweden Follow-Up Study. J Urol. 2010;184(4):1322-7.
348. Vellekoop A, Loeb S, Folkvaljon Y, Stattin P. Population Based Study of Predictors of Adverse Pathology among Candidates for Active Surveillance with Gleason 6 Prostate Cancer. J Urol. 2013.
349. van den Bergh RC, Albertsen PC, Bangma CH, et al. Timing of Curative Treatment for Prostate Cancer: A Systematic Review. Eur Urol. 2013.
350. Klotz L, Vesprini D, Loblaw A. Long term follow-up of a large active surveillance cohort. EAU Stockholm, 29th Annual EAU Congress; Stockholm2014.
351. Popiolek M, Rider JR, Andren O, et al. Natural history of early, localized prostate cancer: a final report from three decades of follow-up. Eur Urol. 2013;63(3):428-35.
352. Glass AS, Cary KC, Cooperberg MR. Risk-based prostate cancer screening: who and how? Current urology reports. 2013;14(3):192-8.
353. Albertsen PC, Moore DF, Shih W, Lin Y, Li H, Lu-Yao GL. Impact of comorbidity on survival among men with localized prostate cancer. J Clin Oncol. 2011;29(10):1335-41.
354. Daskivich TJ, Chamie K, Kwan L, et al. Comorbidity and competing risks for mortality in men with prostate cancer. Cancer. 2011;117(20):4642-50.
355. Daskivich TJ, Fan KH, Koyama T, et al. Effect of age, tumor risk, and comorbidity on competing risks for survival in a U.S. population-based cohort of men with prostate cancer. Ann Intern Med. 2013;158(10):709-17.
356. Lansdorp-Vogelaar I, Gulati R, Mariotto AB, et al. Personalizing age of cancer screening cessation based on comorbid conditions: model estimates of harms and benefits. Ann Intern Med. 2014;161(2):104-12.
357. Zackrisson B, Aus G, Lilja H, Lodding P, Pihl CG, Hugosson J. Follow-up of men with elevated prostate-specific antigen and one set of benign biopsies at prostate cancer screening. Eur Urol. 2003;43(4):327-32.

358. Otto SJ, van der Cruijsen IW, Liem MK, et al. Effective PSA contamination in the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer. Int J Cancer. 2003;105(3):394-9.
359. Bihrmann K, Jensen A, Olsen AH, et al. Performance of systematic and non-systematic ('opportunistic') screening mammography: a comparative study from Denmark. J Med Screen. 2008;15(1):23-6.
360. de Gelder R, Bulliard JL, de Wolf C, et al. Cost-effectiveness of opportunistic versus organised mammography screening in Switzerland. Eur J Cancer. 2009;45(1):127-38.
361. Union CotE. Council recommendation of 2 December 2003 on cancer screening. Off J Eur Union 2003;878.
362. Smailyte G, Aleknaviciene B. Incidence of prostate cancer in Lithuania after introduction of the Early Prostate Cancer Detection Programme. Public health. 2012;126(12):1075-7.
363. van Leeuwen PJ, Roobol MJ, Kranse R, et al. Towards an optimal interval for prostate cancer screening. Eur Urol. 2012;61(1):171-6.
364. Wu GH, Auvinen A, Yen AM, et al. The impact of interscreening interval and age on prostate cancer screening with prostate-specific antigen. Eur Urol. 2012;61(5):1011-8.
365. Stattin P, Carlsson S, Holmstrom B, et al. Prostate cancer mortality in areas with high and low prostate cancer incidence. J Natl Cancer Inst. 2014;106(3):dju007.
366. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA: a cancer journal for clinicians. 2013;63(1):11-30.
367. Drazer MW, Huo D, Schonberg MA, Razmaria A, Eggener SE. Population-based patterns and predictors of prostate-specific antigen screening among older men in the United States. J Clin Oncol. 2011;29(13):1736-43.
368. Vickers A, Carlsson S, Laudone V, Lilja H. It Ain't What You Do, It's the Way You Do It: Five Golden Rules for Transforming Prostate-Specific Antigen Screening. Eur Urol. 2014.
369. Vickers A, Bennette C, Steineck G, et al. Individualized estimation of the benefit of radical prostatectomy from the Scandinavian Prostate Cancer Group randomized trial. Eur Urol. 2012;62(2):204-9.
370. Sun L, Caire AA, Robertson CN, et al. Men older than 70 years have higher risk prostate cancer and poorer survival in the early and late prostate specific antigen eras. J Urol. 2009;182(5):2242-8.
371. Bechis SK, Carroll PR, Cooperberg MR. Impact of age at diagnosis on prostate cancer treatment and survival. J Clin Oncol. 2011;29(2):235-41.
372. Lilja H, Ulmert D, Bjork T, et al. Long-term prediction of prostate cancer up to 25 years before diagnosis of prostate cancer using prostate kallikreins measured at age 44 to 50 years. J Clin Oncol. 2007;25(4):431-6.

373. Carlsson S, Assel M, Sjoberg D, et al. Influence of blood prostate specific antigen levels at age 60 on benefits and harms of prostate cancer screening: population based cohort study. BMJ. 2014;348:g2296.
374. Bul M, van Leeuwen PJ, Zhu X, Schroder FH, Roobol MJ. Prostate cancer incidence and disease-specific survival of men with initial prostate-specific antigen less than 3.0 ng/ml who are participating in ERSPC Rotterdam. Eur Urol. 2011;59(4):498-505.
375. van Leeuwen PJ, Connolly D, Tammela TL, et al. Balancing the harms and benefits of early detection of prostate cancer. Cancer. 2010;116(20):4857-65.
376. Ankerst DP, Hoefler J, Bock S, et al. Prostate Cancer Prevention Trial risk calculator 2.0 for the prediction of low- vs high-grade prostate cancer. Urology. 2014;83(6):1362-7.
377. Ankerst DP, Boeck A, Freedland SJ, et al. Evaluating the PCPT risk calculator in ten international biopsy cohorts: results from the Prostate Biopsy Collaborative Group. World J Urol. 2012;30(2):181-7.
378. SWOP, Prostate Cancer Research Foundation, Rotterdam. ; Available from: http://www.prostatecancer-riskcalculator.com/seven-prostate-cancer-risk-calculators [Accessed 27 Aug 2014]
379. van Vugt HA, Roobol MJ, Kranse R, et al. Prediction of prostate cancer in unscreened men: external validation of a risk calculator. Eur J Cancer. 2011;47(6):903-9.
380. Roobol MJ, Steyerberg EW, Kranse R, et al. A risk-based strategy improves prostate-specific antigen-driven detection of prostate cancer. Eur Urol. 2010;57(1):79-85.
381. Oliveira M, Marques V, Carvalho AP, Santos A. Head-to-head comparison of two online nomograms for prostate biopsy outcome prediction. BJU Int. 2011;107(11):1780-3.
382. Nam RK, Toi A, Klotz LH, et al. Assessing individual risk for prostate cancer. J Clin Oncol. 2007;25(24):3582-8.
383. Nam RK, Kattan MW, Chin JL, et al. Prospective multi-institutional study evaluating the performance of prostate cancer risk calculators. J Clin Oncol. 2011;29(22):2959-64.
384. Thompson IM, Goodman PJ, Tangen CM, et al. The influence of finasteride on the development of prostate cancer. N Engl J Med. 2003;349(3):215-24.
385. Andriole GL, Bostwick DG, Brawley OW, et al. Effect of dutasteride on the risk of prostate cancer. N Engl J Med. 2010;362(13):1192-202.
386. Robinson D, Garmo H, Bill-Axelson A, Mucci L, Holmberg L, Stattin P. Use of 5alpha-reductase inhibitors for lower urinary tract symptoms and risk of prostate cancer in Swedish men: nationwide, population based case-control study. BMJ. 2013;346:f3406.

387. Lucia MS, Epstein JI, Goodman PJ, et al. Finasteride and high-grade prostate cancer in the Prostate Cancer Prevention Trial. J Natl Cancer Inst. 2007;99(18):1375-83.
388. Akre O, Garmo H, Adolfsson J, Lambe M, Bratt O, Stattin P. Mortality among men with locally advanced prostate cancer managed with noncurative intent: a nationwide study in PCBaSe Sweden. Eur Urol. 2011;60(3):554-63.
389. Lowrance WT, Elkin EB, Yee DS, et al. Locally advanced prostate cancer: a population-based study of treatment patterns. BJU Int. 2012;109(9):1309-14.
390. Trinh QD, Bjartell A, Freedland SJ, et al. A systematic review of the volume-outcome relationship for radical prostatectomy. Eur Urol. 2013;64(5):786-98.
391. Savage CJ, Vickers AJ. Low annual caseloads of United States surgeons conducting radical prostatectomy. J Urol. 2009;182(6):2677-9.
392. Fridriksson J, Gunseus K, Stattin P. Information on pros and cons of prostate-specific antigen testing to men prior to blood draw: a study from the National Prostate Cancer Register (NPCR) of Sweden. Scand J Urol Nephrol. 2012;46(5):326-31.
393. Klotz L. Prostate cancer overdiagnosis and overtreatment. Current opinion in endocrinology, diabetes, and obesity. 2013;20(3):204-9.
394. Esserman LJ, Thompson IM, Reid B, et al. Addressing overdiagnosis and overtreatment in cancer: a prescription for change. Lancet Oncol. 2014;15(6):e234-42.
395. Vickers A, Cronin A, Roobol M, et al. Reducing unnecessary biopsy during prostate cancer screening using a four-kallikrein panel: an independent replication. J Clin Oncol. 2010;28(15):2493-8.
396. Cuzick J. Screening for cancer: future potential. Eur J Cancer. 1999;35(5):685-92.
397. Catalona WJ, Partin AW, Sanda MG, et al. A multicenter study of [-2]pro-prostate specific antigen combined with prostate specific antigen and free prostate specific antigen for prostate cancer detection in the 2.0 to 10.0 ng/ml prostate specific antigen range. J Urol. 2011;185(5):1650-5.
398. Cuzick J, Thorat MA, Andriole G, et al. Prevention and early detection of prostate cancer. Lancet Oncol. 2014;15(11):e484-e92.
399. Cary KC, Cooperberg MR. Biomarkers in prostate cancer surveillance and screening: past, present, and future. Therapeutic advances in urology. 2013;5(6):318-29.
400. Roobol MJ, Carlsson SV. Risk stratification in prostate cancer screening. Nature reviews Urology. 2013;10(1):38-48.
401. Thompson J, Lawrentschuk N, Frydenberg M, Thompson L, Stricker P, Usanz. The role of magnetic resonance imaging in the diagnosis and management of prostate cancer. BJU Int. 2013;112 Suppl 2:6-20.
402. Barentsz JO, Richenberg J, Clements R, et al. ESUR prostate MR guidelines 2012. European radiology. 2012;22(4):746-57.

403. Bjurlin MA, Meng X, Le Nobin J, et al. Optimization of Prostate Biopsy: the Role of Magnetic Resonance Imaging Targeted Biopsy in Detection, Localization and Risk Assessment. J Urol. 2014;192(3):648-58.
404. Lee DJ, Ahmed HU, Moore CM, Emberton M, Ehdaie B. Multiparametric magnetic resonance imaging in the management and diagnosis of prostate cancer: current applications and strategies. Current urology reports. 2014;15(3):390.
405. Turkbey B, Mani H, Aras O, et al. Prostate cancer: can multiparametric MR imaging help identify patients who are candidates for active surveillance? Radiology. 2013;268(1):144-52.
406. Stamatakis L, Siddiqui MM, Nix JW, et al. Accuracy of multiparametric magnetic resonance imaging in confirming eligibility for active surveillance for men with prostate cancer. Cancer. 2013;119(18):3359-66.
407. Vargas HA, Akin O, Afaq A, et al. Magnetic resonance imaging for predicting prostate biopsy findings in patients considered for active surveillance of clinically low risk prostate cancer. J Urol. 2012;188(5):1732-8.
408. McNaughton-Collins MF, Barry MJ. One man at a time--resolving the PSA controversy. N Engl J Med. 2011;365(21):1951-3.

ISBN 978-91-628-9224-1
Printed by Ale Tryckteam AB, Bohus