Optimized use of MRI in a PSA-based prostate cancer screening program

Jonas Wallström

Department of Urology and Department of Radiology

Institute of Clinical Sciences

Sahlgrenska Academy, University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2021

"The enigma of arrival"
Guillaume Apollinaire 1880-1918

Cover illustration and author photo: Vide Wallström

SVANENMÄRKET

Trycksak
3041 0234

To my parents
Maureen and
Thomas for a
lifetime of
uninterrupted
love, support,
discussion and
laughter.

Optimized use of MRI in a PSA-based prostate cancer screening program

Jonas Wallström, MD

Department of Urology and Department of Radiology, Institute of Clinical Sciences,
Sahlgrenska Academy,
Sahlgrenska Academy, University of Gothenburg, Sweden

ABSTRACT

The overall aim of this thesis was to optimize different aspects of the use of MRI in screening for prostate cancer. *Paper 1* was based on preoperative MRI in a prostatectomy cohort. *Papers 2-4* were based on data from the ongoing Göteborg Prostate Cancer Screening 2 Trial, a randomized, population-based, long-term trial assessing screening with PSA followed by MRI in men aged 50-61 years in Gothenburg and surrounding municipalities. Biopsies were used as the reference standard.

In *Paper 1* three non-expert readers retrospectively assigned PI-RADSv2 scores in MRI performed at multiple sites. A fair to moderate reader agreement (k-score 0.41) and slightly lower tumor detection (overall 70%) compared to previous reports highlights the importance of a quality assurance program. In *Paper 2* cancer detection with bpMRI was compared with mpMRI in a prospective, paired diagnostic study. Bi-parametric MRI was non-inferior to mpMRI and should be considered the method of choice as it also reduces room turn over time and saves healthy men exposure of gadolinium contrast agents. In *Paper 3* a retrospective analysis of men with peripheral zone PI-RADS 3 lesions was performed. Multivariable regression models were built to assess contrast enhancement, lesion size and, PSA density (PSAD) as predictors of cancer. Only PSAD was strongly correlated to cancer. Selecting men for biopsy based on PSAD could potentially help significantly reduce the number of biopsies but data was not sufficient to establish a clinically reliable threshold. In *Paper 4* PRECISE scores were retrospectively assigned in a 2-year MRI follow-up of men with first-round negative MRI or positive MRI with negative biopsies. Few men were diagnosed with cancer in the second round and most MRI lesions were of stable appearance. This provides important safety data in support of a follow-up interval of at least 2 years.

SAMMANFATTNING PÅ SVENSKA

Prostatacancer är en mångfacetterad sjukdom som innefattar ett brett spektrum av tumörer från mycket långsamväxande till snabbväxande. Målet med att leta efter prostatacancer innan den ger symptom (screening) är att skapa ett nät som fångar aggressiva tumörer i tid för botande behandling. Men om nätet görs för finmaskigt kommer många små och långsamtväxande tumörer utan potential att ge symtom under mannens livstid att utgöra bifångst (överdiagnostik). Genom att införa magnetkameraundersökning (MR) av prostata i ett screeningprogram förbättras möjligheten att välja vilka män som behöver utredas vidare med vävnadsprover. Den här avhandlingen behandlar olika aspekter av hur användningen av magnetkameraundersökningen kan anpassas för screening.

Första delarbetet baseras på en grupp av 97 män undersökta med MR före operation på ett mindre sjukhus. Resterande delarbeten baseras på den sedan 2015 pågående Göteborg 2-studien inom vilken män i åldern 50–61 år boende i Göteborg med kranskommuner lottas till antingen kontrollgrupp eller screening. Screeninggruppen erbjuds PSA-prov och därefter MR-undersökning vid förhöjt PSA-prov. Män som inte diagnostiserades med cancer i första omgången återinbjuds för upprepad screening. Under första studieomgången som slutfördes 2020 lottades mer än 38 000 män till screening-gruppen och fler än 2000 MR-undersökningar genomfördes.

Sammanfattning av resultat/slutsatser: **1)** MR-protokoll och bedömning av undersökningarna behöver kvalitetssäkras för att nå upp till de goda resultat som redovisats i tidigare studier; **2)** Ett MR-protokoll utan kontrastmedel rekommenderas i screening eftersom det ger lika god tumördetektion som undersökning med kontrastmedel: **3)** Cancerförekomsten var låg i den totalt sett relativt stora andelen av män med osäkra fynd på MR. PSA-densitet kan vara en värdefull parameter för att bättre välja vilka män som behöver undersökas vidare med vävnadsprov men urvalet var för litet för att etablera ett gränsvärde **4)** Mycket få män med normala fynd på MR och/eller normalt vävnadsprov diagnosticeras med cancer vid uppföljning efter 2 år. Män med klart misstänkta fynd på MR behöver dock följas upp tidigare eftersom ett litet antal högriskcancrar missades vid vävnadsprovtagning.

LIST OF PAPERS

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

I. Kohestani K, **Wallström, J**, Dehlfors N, Sponga O, Månsson M, Josefsson A, Carlsson S, Hellström M, Hugosson J. *Performance and inter-observer variability of MRI (PI-RADS version 2) outside high-volume centers.* Scandinavian Journal of Urology 2019; 53:5: 304-311, DOI: 10.1080/21681805.2019.1675757

II. **Wallström, J**, Geterud K, Kohestani K, Maier S, Månsson M, Pihl C-G, Socratous A, Arnsrud-Godtman R, Hellström M, Hugosson J. *Bi- or multiparametric MRI in a sequential screening program for prostate cancer with PSA followed by MRI? Results from the Göteborg Prostate Cancer Screening 2 Trial.* European Radiology 2021, DOI:10.1007/s00330-021-07907-9

III. **Wallström, J**, Månsson M, Axcrona U, Egevad L, Geterud K, Kohestani K, Maier S, Pihl C-G, Socratous A, Arnsrud-Godtman R, Hellström M, Hugosson J. *Evaluation of contrast enhancement, lesion area and PSA density in selecting men with PI-RADS 3 lesions for biopsy. Results from the Göteborg Prostate Cancer Screening 2 Trial.* In manuscript.

IV. **Wallström, J**, Geterud K, Kohestani K, Maier S, Pihl C-G, Socratous A, Stranne J, Arnsrud-Godtman R, Månsson M, Hellström M, Hugosson J. *Outcomes of repeated MRI after 2 years. Results from the Göteborg Prostate Cancer Screening 2 Trial.* Submitted.

CONTENT

Abbreviations

MRI	Magnetic Resonance Imaging
bpMRI	Bi-parametric MRI
mpMRI	Multi-parametric MRI
DWI	Diffusion Weighted Imaging
DCE	Dynamic Contrast Enhanced
EPE	Extraprostatic Extension
PI-RADS	Prostate Imaging Reporting and Data System
GS	Gleason Score
PC	Prostate Cancer
PSA	Prostate Specific Antigen
PSAD	Prostate Specific Antigen Density
SBX	Systematic Biopsies
TBX	Targeted Biopsies
TPM	Trans-perineal Prostate Mapping
TRUS	Transrectal Ultrasound

DEFINITIONS IN SHORT

Prostate MRI pathway   Prebiopsy prostate MRI with targeted biopsies in case of findings at MRI suggestive of cancer

Cancer screening   Testing for disease when there are no symptoms

1   INTRODUCTION

For the past years, I have had the privilege to work in the dynamic borderland between radiology and urology, with the introduction of magnetic resonance imaging (MRI) in the diagnostic pathway of the most common malignancy in males, prostate cancer. The prostate MRI pathway has dramatically changed the management of men under clinical suspicion of prostate cancer with evidence that more clinically significant cancer, and less indolent cancer is detected with MRI followed by targeted biopsies compared to upfront systematic biopsies. Screening for prostate cancer, however, remains an unsolved question mainly because of major concerns with overdiagnosis of indolent cancer. In the Göteborg 2 Trial, the effects of introducing the prostate MRI pathway in a PSA-based screening program for prostate cancer is studied.

Prostate cancer is a disease with a multitude of different faces reflecting the diverse tumor biology and aggressiveness of individual cancers. At the one end of the spectrum, highly aggressive prostate cancer can be compared to a tiger, which already at the time of diagnosis is a predator that will spread throughout the body, dramatically reduce quality of life and ultimately result in premature demise. At the other end of the spectrum, indolent prostate cancer is more akin to a domesticated cat, that will not interfere with life in any noticeable way, and should best be left undiagnosed. But in the middle ground, between ferocious tigers and pet cats, there is an interesting group of felines - representing cancers that could be cured if timely diagnosed.

In another analogy, which I have borrowed from Gilbert Welch [1], screening for cancer can be compared to a farmer contemplating constructing a fence for his - somewhat haphazard collection of animals - turtles, birds, and hares. A fence should contain most of the hares (curable cancers with malignant potential) but fencing of turtles (indolent cancers) would be a waste of time and fencing of birds (cancers spread at presentation) would not be possible.

This thesis deals with a new component of the fence – prostate MRI - and how it can be optimized for use in a PSA-based screening program.

1.1 PROSTATE MRI AND PROTOCOLS

In 1982 Steyn and Smith published the first report on prostate MRI performed in 25 men using a prototype machine [2]. Four years later, in 1986, Hricak et al published a report on normal prostate anatomy at 0.35 and 1.5 T MRI with depiction of the zonal anatomy using T2-weighted imaging [3].

For anatomic detail, T2-weighted imaging remains the most important pulse sequence in the protocol. The zonal anatomy of the prostate is nicely visualized, separating peripheral zone from transition zone and true central zone. The prostate "pseudo-capsule" is delineated as a dark signal band around the prostate joining the fibromuscular stroma in the anterior part. The seminal vesicles, the prostatic urethra, and vas deferens are also depicted as well as the surrounding anatomy including the bladder neck and lower sphincter, all important structures to evaluate before curative treatment with surgery or irradiation (Fig. 1).

Figure 1. 3 Tesla T2-weighted anatomic prostate imaging. Top left: Axial, 3 mm slice thickness obtained at large field of view to cover pelvic nodes. Top right: Oblique axial section, 1.5 mm slice thickness, shows bright signal in the peripheral zone with some thin streaks and surrounding low signal "capsule" ventrally blending with the fibromuscular stroma. Centrally, the transition zone with multiple benign appearing nodules is seen. Bottom left: Sagittal section, 1.5 mm slice thickness. Bottom right: Coronal section, 1.5 mm slice thickness.
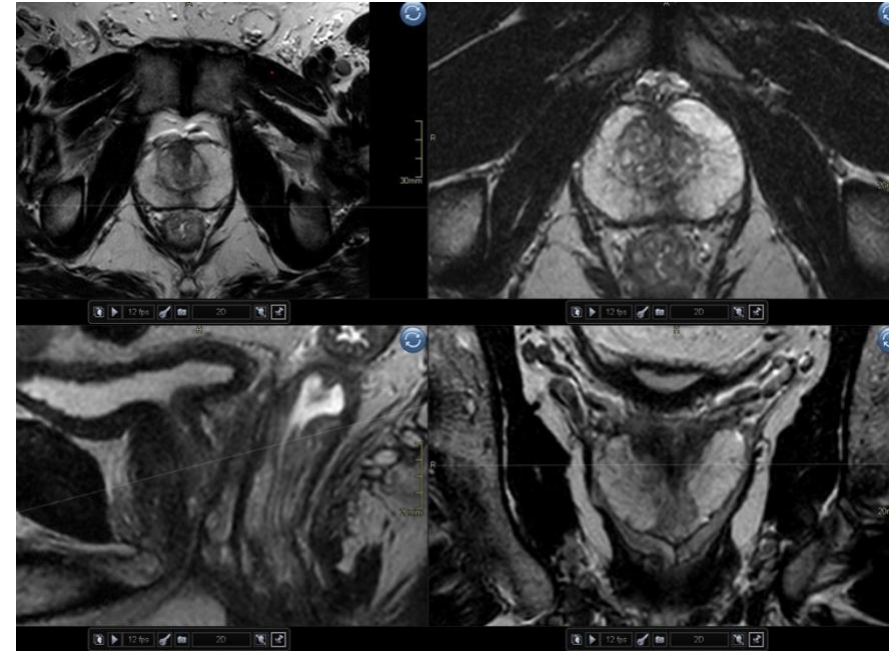
Initially much interest was focused on tumor staging using the anatomic detail afforded by T2-weighted images. In addition, T1-weighted images were acquired, important for assessment of both post-biopsy hemorrhage and bone marrow lesions. The next major addition to the protocol was contrast-enhanced T1-weighted imaging. In 1995 the first published report on the use of dynamic contrast enhanced imaging (DCE) concluded that tumor margins were better depicted in relation to the prostatic capsule and seminal vesicles [4].

To set the stage for a shift towards using prostate MRI for early tumor detection additional developments of the protocol were needed, in particular a pulse sequence that would enable more accurate differentiation of malignant from

benign prostatic tissue [5]. In 2002 Issa published the first report on prostatic in vivo measurement of the Apparent Diffusion Coefficient (ADC) with diffusion-weighted imaging (DWI) of the prostate [6]. With the addition of DWI/ADC to prostate MRI, all three – at present time routinely used - components of the multi-parametric protocol were in place.

Spectroscopy was introduced around the same time [7] as DWI but DWI proved to be a more robust approach and spectroscopy is now seldom used.

In current protocols DWI, is the workhorse for tumor detection, particularly in the peripheral zone, where the majority of tumors are located. Most malignant lesions are histologically dense compared to the normal gland, restricting the free diffusion of water molecules and hence producing an increased signal on high b-value DWI and low signal on apparent diffusion coefficient (ADC) maps [8]. However, a certain overlap in imaging features of histologically benign findings such as fibrosis/inflammation and malignant findings exists and can sometimes make differentiation impossible [9].

The vascularity of the tumor is assessed with DCE [10]. A gadolinium-based contrast medium is injected intravenously and imaging is performed for at least 2 minutes with a temporal resolution of <15 seconds. A highly neo-vascularized tumor will both enhance and wash out early compared to normal glandular tissue [10]. In the beginning, much effort was put into classifying enhancement curve types based on wash in and wash out of contrast agents. However, many malignant tumors do not clearly exhibit these characteristics and attention has now shifted to "eyeballing" lesions for early or contemporaneous enhancement, scored as either "positive" or "negative" [11]. As benign hyperplastic nodules avidly enhance, DCE is generally not considered of value for tumor assessment in the transition zone [12].

It has been observed that Gadolinium contrast agents, in particular linear molecules, can be retained in the brain and other tissues but no adverse effects have been proven after decades of clinical use. Currently favored macro-cyclic contrast agents however have a very high safety profile [13].

Two major protocols are currently used:

Multi-parametric MRI (mpMRI) including T2-weighted imaging, DWI, and DCE. In the PI-RADSv2.1 guideline [14] mpMRI is the default protocol. Including DCE can increase lesions conspicuity and also provides a safety net in case of low-quality DWI due to artifacts as DCE is less susceptible to artifacts compared to DWI.

Bi-parametric MRI (bpMRI) including only T2-weighted imaging and DWI. Abandoning DCE makes imaging non-invasive and is beneficial particularly in terms of reduced cost and improved logistics. Several studies have reported comparable diagnostic accuracy with bpMRI and mpMRI [15-21].

1.1.1 PI-RADS

In an effort to standardize the performance and reading of prostate MRI the European Society of Urogenital Radiology (ESUR) drafted guidelines known as Prostate Imaging-Reporting and Data System in 2012 [22]. Three years later, in 2015, the guidelines were updated to version 2 in collaboration with the American College of Radiology (ACR) [23-25].

A minor update to PI-RADSv2.1 was issued in 2019 including updated scoring of transition zone nodules and clarification of interpretation criteria for DWI and DCE [14]. A statement on bi-parametric MRI was issued in the 2.1 update and the considerations of the PI-RADS committee on bpMRI in biopsy naïve men were further developed in a narrative review in 2020 discussing the possible use of bpMRI in low-risk populations and in settings with only experienced readers [26].

In PI-RADSv2.1 the likelihood that a lesion corresponds to prostate cancer with a *Gleason score of ≥7*, and/or *tumor volume >0.5 ml* and/or *extra-prostatic extension (EPE)* is scored on a five-point scale [27]. The likelihoods are defined as *very high, high, intermediate, low,* and *very low* for overall assessment categories 5, 4, 3, 2, and 1 respectively. The guidelines broadly recommend biopsy of lesions scored PI-RADS ≥4 but do not offer a comprehensive system of biopsy recommendations.

A weak correlation between Gleason score and MR imaging signal intensity exists. In the peripheral zone, the correlation is stronger for ADC/DWI and in the transition zone, the correlation is stronger for T2-weighted imaging [28]. Thus, overall scoring is based on prostate zonal anatomy and the concept of dominant pulse sequences.
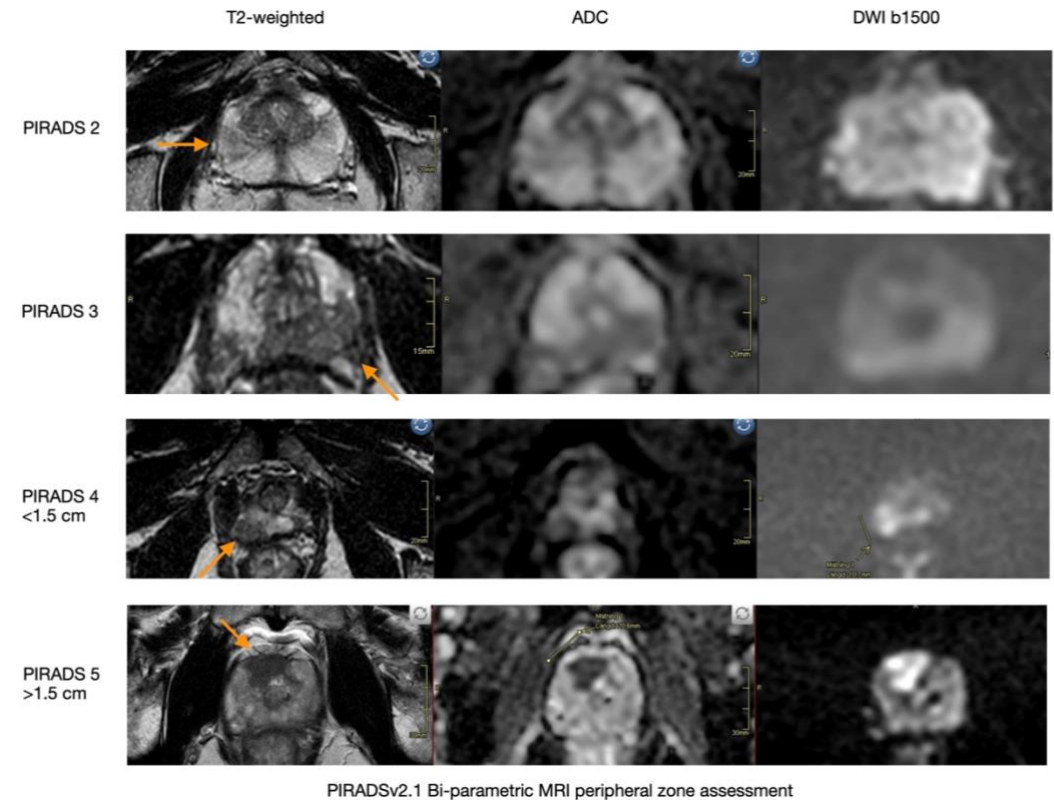
In the *peripheral zone,* the dominant sequence is DWI. The overall assessment category is the same as the DWI score except if the DWI score is indeterminate (DWI=3) and DCE is positive in which case the overall assessment is "DCE-upgraded" to PI-RADS=4.

In the *transition zone,* the dominant sequence is T2W. The overall assessment category is usually the same as for T2W but findings on DWI upgrade the overall score in case of "atypical" nodules or indeterminate T2W and markedly restricted DWI with size >1.5 cm.

The PI-RADS assessment algorithm is illustrated in Figure 4 and pictorial examples are provided in Figures 2 and 3.
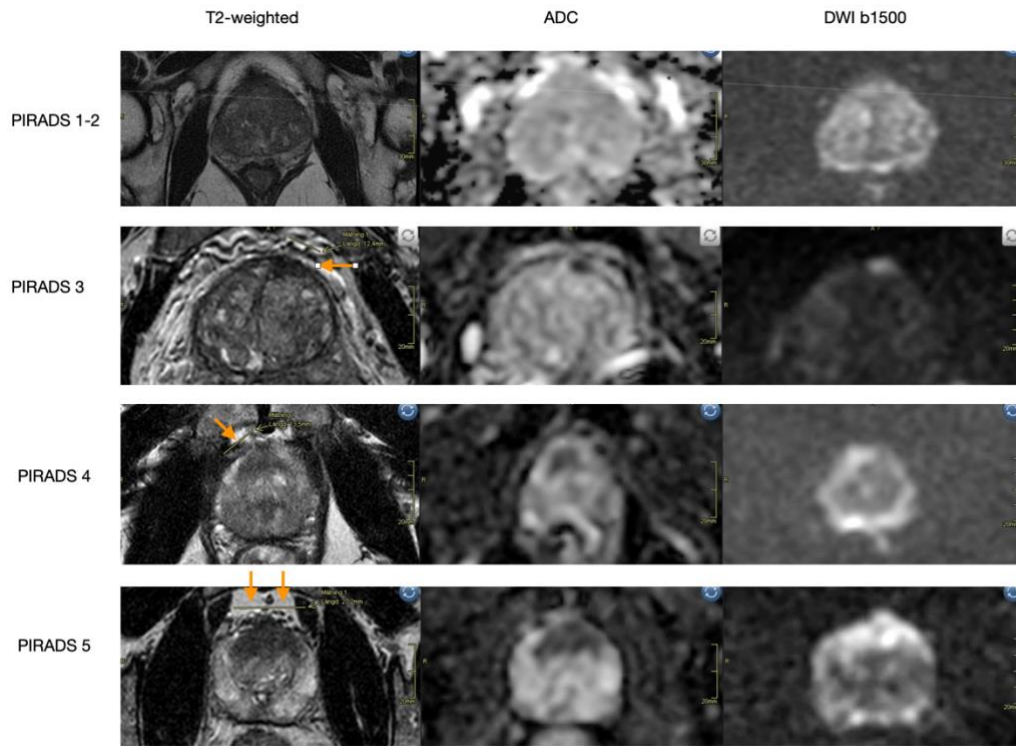
In several studies involving validation of PI-RADSv2, it was observed that the biopsy yield of significant cancer increases with increasing PI-RADS overall assessment category [29-31].

Figure 2. PI-RADS peripheral zone examples

PIRADSv2.1 Bi-parametric MRI peripheral zone assessment

From top to bottom: PI-RADS 2: Lesion in right lateral sector, wedge shaped area, no restricted DWI; PI-RADS 3: Lesion in left posterolateral sector, diffuse lesion with moderately restricted DWI; PI-RADS 4: Lesion in right posterolateral sector, 10 mm, focal lesion with markedly restricted DWI; PI-RADS 5: Lesion in right anterior sector, 20 mm, focal lesion with markedly restricted DWI.

Figure 3. PI-RADS transition zone examples

From top to bottom: PI-RADS 1-2: Multiple, well circumscribed benign hypertrophic nodules; PI-RADS 3: Lesion in left anterior sector, 10 mm nodule with non-circumscribed margins and restricted DWI; PI-RADS 4: Lesion in right anterior sector, 12 mm, lenticular lesions with markedly restricted DWI: PI-RADS 5: Lesion in anterior sectors, 20 mm, lenticular lesion with markedly restricted DWI.

Figure 4. PI-RADS scoring of lesions in the peripheral zone. DWI is the dominant pulse sequence for overall scoring. Positive DCE only upgrades in case of DWI = 3; DCE-upgraded PI-RADS 4 (3+1).

DWI	T2W	DCE	PI-RADS
1	1-5	-/+	1
2	1-5	-/+	2
3	1-5	-	3
3	1-5	+	4 (3+1)
4	1-5	-/+	4
5	1-5	-/+	5

Figure 5. PI-RADS scoring of lesions in the transition zone. T2W is the dominant pulse sequence for overall scoring. DWI ≥4 upgrades in case of T2W = 2 or 3. No upgrading is performed based on DCE.

T2W	DWI	DCE	PI-RADS
1	1-5	-/+	1
2	≤3	-/+	2
2 "Atypical nodule "	≥4	-/+	3
3	≥4	-/+	3
3	5	-/+	4
4	1-5	-/+	4
5	1-5	-/+	5

1.1.2   PRECISE

In 2017 a multidisciplinary working group put together a set of criteria for reporting MRI-based active surveillance studies known as the PRECISE recommendations [32]. To facilitate the assessment of tumor progression compared to previous MRI examinations, it was recommended that the likelihood of significant tumor progression should be assessed and reported with a five-point Likert scale in combination with a description of which parameters were changed. Progression was defined as PRECISE >4 and regression as PRECISE <3 (Fig. 6). No specific recommendations on thresholds for change in size or conspicuity were given due to lack of robust data to support such thresholds.

Figure 6. PRECISE scores

	Definition	Example
PRECISE1	Complete regression	Complete resolution
PRECISE2	Partial regression	Reduction in size or conspicuity
PRECISE3	Stable imaging appearance	No change in size or conspicuity
PRECISE4	Progression	Increase in size or conspicuity/increased PI-RADS score and/or new lesion
PRECISE5	Tumor stage progression	Progressions with extracapsular extension and/or seminal vesicle invasion

1.1.3   PATHOLOGY

The modern foundation of grading Prostatic carcinoma was laid by Donald F. Gleason in a seminal paper published in 1966 [33]. Gleason identified five histological patterns (Gleason pattern 1-5) associated with increased tumor aggressiveness (Fig. 7). The pathology report includes two patterns that are summed, Gleason score (GS), for example, 3+3 = 6 in case of low-risk cancer. Initially, the sum was made up of the two most common patterns. In 2005 the International Society of Urological Pathology (ISUP) consensus meeting changed the definition to always include the highest grade pattern when assessing biopsy cores, thus adding the most common pattern and the highest grade pattern in any of the cores [34]. Currently, it has become common in the United States to report only the core with the highest scores, but in Europe, it is more common to report the highest grade and the most usual grade.

In 2014 a new ISUP consensus meeting was held introducing the concept of grade groups for reporting biopsy findings, also known as ISUP grade groups [35]. The ISUP grade groups are numbered 1-5, ISUP 1 corresponding to GS 3+3, ISUP 2 corresponding to GS 3+4, ISUP 3 corresponding to GS 4+3, ISUP 4 corresponding to any Gleason sum 8, and ISUP 5 to any Gleason sum 9-10. Although easier to communicate with patients and incorporated in WHO classification of tumors of the urogenital system and male genital organs in 2015 the system of grade groups has been criticized for not taking into account clinically relevant parameters such as the percentage of Gleason 4 pattern or the increased risk of recurrence with any Gleason 5 component reported in some studies [36].

To pathologically define clinically significant prostate cancer is not easy. It is known from autopsy studies that the prevalence of prostate cancer increases by approximately 10 percentage points with every decade of life from the age of 20 and onward [37]. It is noteworthy that a large proportion of these cancers will never become clinically relevant and, as previously mentioned in the Introduction section, PSA-screening followed by systematic biopsies resulted in substantial over-diagnosis of indolent cancers.

In a study by Stamey et al in 1993 [38] the lifetime probability of clinically significant prostate cancer was estimated to 8% based on data from the US National Cancer Institute and it was shown that in men who underwent cystoprostatectomy the largest 8% of detected prostate cancers had a tumor volume of larger than 0.5 mL. It was also concluded that the 80% of detected prostate cancers with a tumor volume below 0.5 mL were unlikely to reach clinical significance considering the long doubling time of prostate cancer. In

addition, in a study including over 14000 radical prostatectomies with GS≤6 cancer no single case of lymph node metastases was identified [39] supporting a low biological potential of GS≤6 tumors to metastasize.

The Epstein criteria were developed to select men with low-risk cancer for management with active surveillance [40, 41]. The criteria were based on 6 systematic biopsy cores and later included in the US National Comprehensive Cancer Network (NCCN) guidelines criteria of "very-low-risk cancer" where they are defined as non-palpable cancers with a maximum of two biopsy cores with no more than 50% cancer with GS≤6 and PSA<10 ng/ml (PSAD <0.15 ng/ml^2) [42].

This means that a great deal of our understanding of the clinical importance of low, middle, and high-risk prostate cancer as defined by the Gleason score is based on data from systematic biopsies which is problematic when redefining criteria to fit with the prostate MRI pathway.

For the past decades, there has been continued work to define pathological criteria for clinically significant versus clinically insignificant prostate cancer. Furthermore, with the move towards early biopsy diagnosis criteria from prostatectomy studies have to be translated first to systematic biopsies and now to targeted biopsies in the MRI pathway.

Figure 7. Gleason's five patterns: 1 = Small, uniform glands; 2 = Increased stroma between glands; 3 = Infiltrative margins; 4 = Neoplastic glands forming irregular masses; 5 = Glands only formed occasionally (based on reference #33).
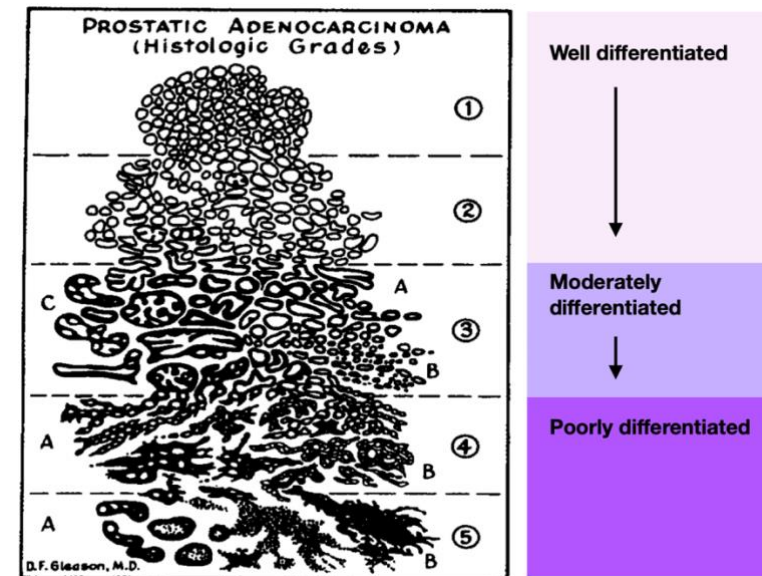
1.1.4 TARGETED PROSTATE BIOPSIES

Targeted prostate biopsies can either be performed with ultrasound guidance or in the MRI room as "in gantry" biopsies.

The most common approach is to use ultrasound guidance. Transrectal ultrasound-guided biopsies (TRUS) are performed with a rectal probe. Trans-perineal biopsies are performed with a linear surface probe.

If ultrasound guidance is used the biopsy needles can either be placed by cognitive fusions or by software-assisted fusion. In cognitive fusion, the operator first localizes the lesion on the MRI images, and then cognitively fuses the image with the ultrasound image at biopsy. In software assisted fusion the lesions are first contoured on the MRI images, and then fused with the ultrasound images in real-time. If MRI guidance is used the needles are placed while the patient is inside the MRI gantry.

The number of targeted cores per lesion is usually 1-2 cores if MRI in gantry biopsies are performed, and 3-4 cores, if ultrasound-guided biopsies are performed.

Although a systematic review in 2016 reported a possible advantage for detection of significant PC using software-assisted fusion - or MRI in gantry biopsies - over cognitive fusion biopsies [43], both a retrospective cohort study performed in 2017 [44] and a recent randomized trial showed no significant difference in detection rates with all techniques [45]. In a study comparing the detection rate of significant PC (≥ISUP2) with cognitive targeted biopsies versus Trans-perineal prostate mapping (TPM) biopsies no statistically significant difference was found between the methods [46].

1.1.5 THE PROSTATE MRI PATHWAY

For the last decade, there has been a steadily increasing interest in performing pre-biopsy MRI and targeted biopsies [46-48].

The prostate MRI pathway is a chain of diagnostic events including MRI, targeted biopsies, and histological assessment of biopsy cores. In the MRI pathway men with negative MRI are not biopsied. PSA-levels and other pre-MRI tests such as STHLM-3 or 4KS could also be considered part of the pathway by modifying the pre-MRI probability of detecting PC.

In the much-cited multicenter "PRECISION"-trial 500 biopsy naïve men under clinical suspicion of prostate cancer were randomized to either MRI followed by targeted biopsies in case of PI-RADS ≥3 findings, or upfront systematic biopsies without preceding MRI. MRI with targeted biopsies detected more significant cancer (38% versus 26 %) and less insignificant cancer (9% versus 22%) compared to systematic biopsies. In the MRI group, 28% of the men had a negative MRI examination and avoided biopsies [49].

Other randomized controlled trials have however not reported a significant advantage of targeted biopsies over systematic biopsies in detecting ISUP ≥2 PC in men not previously biopsied but all studies show that overdiagnosis of low-risk PC is reduced with the MRI pathway [50-52]. The EAU-ESUR-guidelines on prostate cancer recommends that MRI should be performed before biopsy in men under clinical suspicion of prostate cancer but not as an initial screening tool [53].

In a recent Cochrane review assessing prostate MRI with template-guided biopsies as the reference standard, the pooled sensitivity and specificity of the MRI pathway compared to systematic biopsies for clinically significant prostate cancer was 0.72 and 0.96 compared to 0.63 and 1.0 respectively. The review concluded that: "the MRI pathway has the most favorable diagnostic accuracy in significant prostate cancer detection" although the level of evidence was considered low due to weaknesses and inconsistencies in the included studies [54].
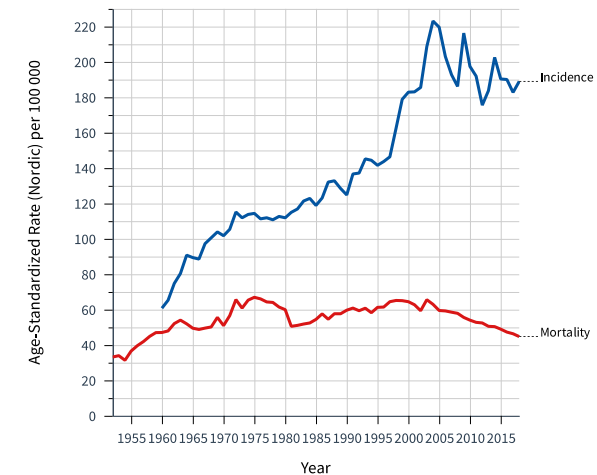
1.2   SCREENING FOR PROSTATE CANCER

1.2.1   PROSTATE CANCER EPIDEMIOLOGY

Clinical prostate cancer (PC) is seldom diagnosed before the age of 50 years. In 2018 PC was the cancer with the highest incidence among men in Europe, and the third cause of cancer death after lung- and colorectal cancer [55].

In the Nordic countries, the age-standardized incidence of PC almost doubled in the early era of PSA-testing 1990-2004 but mortality was largely unaffected (Fig. 8). However, for the last two decades, the mortality rates have been steadily decreasing in all the Nordic countries [56].

In 2016 more than 10,000 Swedish men were diagnosed with PC and over 100,000 men lived with the diagnosis. It is especially noteworthy that the mortality rate has decreased by almost 35% in Swedish men under 75 years of age since 2005 [57]. The total number of deaths has however remained almost constant, around 2400 per year, reflecting an increasing proportion of elderly men in the population [58].

Figure 8. Age standardized prostate cancer incidence and mortality in Sweden 1955-2016. Data from Nordcan[59, 60] .

1.2.2 HARMS AND BENEFITS OF PSA-BASED SCREENING FOR PC

In the mid-1990s, during the early days of PSA-testing, the incidence of prostate cancer dramatically increased, but the mortality was largely unchanged, reflecting the detection of indolent cancer, i.e., overdiagnosis [61-63]. Overdiagnosis was mainly driven by the practice of systematic biopsy sampling of the prostate with multiple cores directed at the dorsal parts of the gland. Initially, six-core "sextant" biopsies were standard but later on the number of cores was increased to twelve.

With systematic biopsies, large tumors are usually covered, but small tumors can easily be missed, and tumors in the ventral aspect of the gland are not included in the biopsy scheme [64, 65]. Furthermore, the risk of incidentally hitting a small indolent cancer is rather large. Apart from discomfort and the risk of overdiagnosis transrectal biopsies carry at least a 2-3% risk of serious infection (requiring the patient to be hospitalized) despite the use of prophylactic antibiotics [66]. Furthermore, many common pathogens are increasingly resistant to antibiotics which discourages prophylactic use.

Spurred by the widespread implementation of PSA-testing large-scale prostate cancer screening studies were started in both Europe and the US. The European Randomized Study of Screening for Prostate Cancer (ERSPC) was a multicenter study including 162000 men aged 55-59 years, that were randomized to either a control group or PSA-testing, with a repeat interval of 2 or 4 years. The study was limited by a significant crossover between study arms with 23-40% opportunistic PSA-testing in the control arm [67]. At 16-year follow up the number needed to screen (NNS) to prevent one prostate cancer death was 570 and the number needed to diagnose (NND) was 18. Thus 18 men had to be diagnosed to prevent one death from prostate cancer. The prostate cancer specific mortality was reduced by 20% in the screening arm compared to the control arm. The absolute reduction in mortality was however only 0.17 percentage points (0.89% controls, 0.72% screening) [68]. All-cause mortality was not significantly reduced at an earlier 9 year follow-up of the same study [67]. In perspective, the benefits offered by PSA-screening could be compared to the reported 0.15% absolute reduction of all-cause mortality by bike-commuting to work, compared to a non-active lifestyle, in a prospective cohort study including 263000 participants followed during 5 years [69].

In its most recent evaluation of PSA-based screening the Swedish National Board of Health and Welfare [70] - in line with European and US guidelines –concluded that the benefits of screening did not clearly outweigh the risks of overdiagnosis and overtreatment on a population level. However, the Ministry of Health and Social affairs the same year commissioned an inquiry by the Regional Cancer Center Boards on the standardization of PSA-testing [71]. The aim was to reduce inequalities of PSA-testing due to socioeconomic and geographical factors, and to offer information and testing to men in the relevant age group of 50-74 years, and to discourage testing in men with a life expectancy of less than 10 years. A second aim was to identify knowledge gaps in complementary diagnostic testing, including imaging such as MRI [71].

The ongoing Göteborg Prostate Cancer Screening 2 (G2) Trial started in 2015 and is part of an effort to answer the question of whether MRI can balance the benefits and harms of prostate cancer screening.

2 AIMS

The overall aim of this thesis was to optimize different aspects of the use of prostate MRI in a PSA-based screening program for prostate cancer.

The specific aims of each paper were:

I. To evaluate prostate cancer detection rate and inter-observer agreement in PI-RADS scoring between readers of different levels of experience outside high-volume centers.

II. To compare cancer detection rates of bi-parametric prostate MRI with multi-parametric MRI in the Göteborg Prostate Cancer Screening 2 Trial.

III. To assess the frequency of cancer in indeterminate peripheral zone (PI-RADS 3) lesions in the Göteborg Prostate Cancer Screening 2 Trial and to evaluate contrast enhancement, lesion size, and PSA density as predictors of detecting significant prostate cancer in these lesions.

IV. To study the risk of prostate cancer during a 2 year MRI follow-up in the Göteborg Prostate Cancer Screening 2 Trial in men with MRI lesions not proven to be cancer in the first round.

Figure 9. Example cases from Paper II: top left T2 axial, top right DCE, bottom left ADC and bottom right DWI. Top panel: PI-RADS 4 (3+1) only scored with mpMRI, low risk PC at biopsy. Bottom panel: False positive mpMRI PI-RADS 4 (3+1).
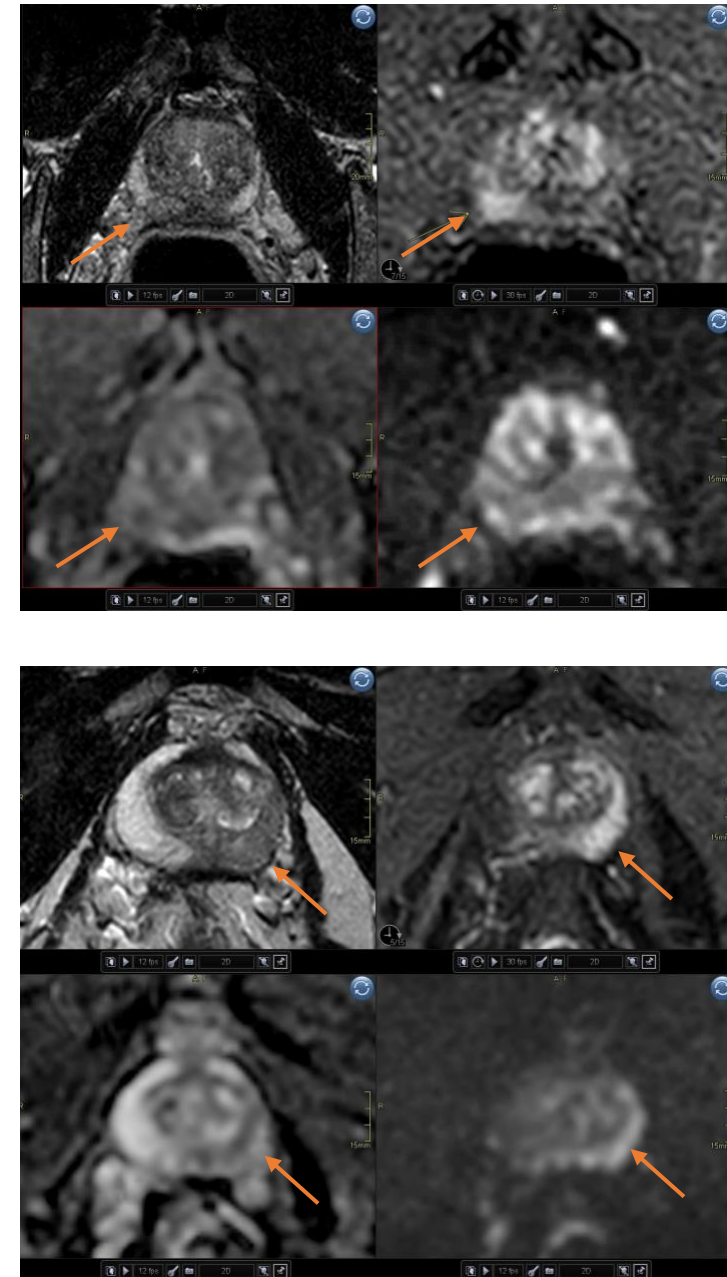
2.1 DESCRIPTION OF STUDY POPULATIONS

Paper I:

The patient cohort consisted of a series of 97 consecutive men diagnosed with prostate cancer, who were selected for prostatectomy at a single hospital, and who had been examined with prostate MRI before surgery. The median age of included men was 61 years. The median PSA at diagnosis was 6.4 ng/mL and the median PSA density based on TRUS volume was 0.21 ng/ml². At prostatectomy, the majority of men (77%) had a tumor size >10 mm and a Gleason score of ≥7 (77%). The pathological stage was pT2 in 63% and pT3 in 37% of the specimens. Positive surgical margins were found in 19% of the specimens.

Papers II-IV:

Inclusion criteria in the G2-trial were men aged 50-61 years, alive at randomization date with a registered address in the county of Gothenburg, or any of six specified surrounding municipalities. Men with a previous diagnosis of prostate cancer, or who emigrated, or died in the period between randomization and update of the total population registry were excluded.

Eligible men were randomized from the population registry and allocated 2:1 to PSA-screening or control group.

Participating men were further randomized to one out of three study arms and in case of PSA-levels elevated above the cut-off (PSA>3 ng/mL in Arms 1-2 and PSA >1.8 ng/mL in Arm 3) they were further invited to MRI.

In *Paper II*, consecutive men in the G2-trial who were examined with MRI between 1 March 2019 and 1 June 2020 were included.

In *Paper III*, consecutive men in the G2-trial who were examined with MRI up to 20 October 2020 were included.

In *Paper IV*, consecutive men in the G2-trial who were examined with a second screening round follow-up MRI up to 30 June 2020 were included.

2.2 CLINICAL INVESTIGATIONS

Paper I:

All prostatectomies were performed by a single highly experienced prostatic surgeon. The operation techniques used were retropubic radical prostatectomy and, in a minority of cases, robot-assisted laparoscopic prostatectomy. Whole-mount prostatectomy specimens were prepared and evaluated according to clinical routines with assessment of Gleason score (GS), pathological Tumor stage, and surgical margins. Pathology reports and scanned whole mount slides were used for correlation with MRI findings.

Papers II-IV:

The screening algorithm in the G2 trial draws a random sample of men aged 50-61 years, living in Gothenburg or surrounding municipalities, from the population registry and allocates them to 2:1 to screening or control group. Men in the screening group are further randomized to one out of three study arms. Arm 1 is the reference arm and includes both systematic and targeted biopsies for men with PSA levels > 3 ng/mL regardless of MRI findings. Arm 2 only includes targeted biopsies in the case of positive MRI for men with PSA levels > 3 ng/mL. Arm 3 only includes targeted biopsies in the case of positive MRI, but in this case for men with PSA-levels > 1.8 ng/mL [72].

Men with a PSA level below the study arm cut off, or incomplete screening (no MRI or no biopsy), are re-invited at pre-specified intervals ranging from 2-8 years. Invitation to screening is stopped at age 62-75 depending on prespecified PSA levels or at age 70 in the case of non-responders.

Men who complete a screening round without being diagnosed with PC are re-invited after 2 years.

2.3 MRI TECHNICAL CONSIDERATIONS AND COMMENTS

Paper I:

The heterogeneity of MRI protocols in *Paper I* reflects the early days of prostate MRI in Sweden, before widespread adoption and standardization in national guidelines. In total imaging was performed at 16 different sites using a variety of MRI platforms, all 1.5T except one 3T MRI at an external site. Protocols were not standardized across sites, for instance, most external sites did not use dynamic contrast-enhanced imaging (DCE), and the b-values used for diffusion-weighted imaging (DWI) and calculation of apparent diffusion coefficient (ADC) maps were not the same.

The majority of MRI examinations were however performed at the main study site using a 1.5 T MRI (GE Medical Systems Signa) with a phased-array pelvic coil. The protocol was updated to multi-parametric MRI including dynamic contrast-enhanced images (DCE) after the publication of PI-RADSv1 in 2012 [22], with the bulk of included MRI examinations (56 out of 66) performed with mpMRI. For DCE 0.1 mmol/kg gadoterate meglumine (Dotarem, Gothia Medical) was administered intravenously using a power injector. The temporal resolution was 15 s.

For DWI, the main site protocol included a high b-value set of b=1500 s/mm^2. To optimize scan time only two b-value sets were acquired, b=0 s/mm^2 and b=1500 s/mm^2, and these were used for calculating ADC maps. Calculating extrapolated high b-values was not an option, since the scanner lacked software support for performing such processing.

T2-weighted images were acquired in three planes including axial 3 mm slices as recommended in the PI-RADS guidelines.

Regarding external sites, only two sites used multi-parametric MRI. All external sites included DWI with a high b-value set of ≥1000 s/mm^2. T2-weighted images included at least two planes, including an axial plane.

Both the first and second versions of the PI-RADS guidelines have set minimum requirements for imaging protocols [14, 22], however, historically full protocol adherence has been low in prostate MRI studies [73]. It can be argued that certain aspects of protocol parameters are more important than others, for instance, the inclusion of a high b-value set of b≥1400 is considered of paramount importance for tumor detection in combination with ADC. However, the importance of adhering to the recommendation to use a low b-value that is higher than 0 and a high b-value of b≤1000 s/mm^2 or less for calculation of the ADC map (to avoid significant departure from mono-exponential diffusion) can be questioned when visually assessing ADC since it can be shown that the contrast ratio between tumor and normal prostate is maintained, although the measured ADC will be lower for both tumor and normal prostate [74]. Low contrast to-noise-ratio (CNR) is generally a bigger problem with loss of anatomic definition and tumor visibility at high b-values if the MRI protocol is not optimized.

It can be speculated that the reported tumor sensitivity in this study was negatively affected by both lack of b=1500 s/mm^2 at some external sites and problems with sub-optimal DWI contrast-to-noise ratio in some of the included examinations.

The importance of DCE is controversial [26]; it has been argued that DCE acts as a "safety net" in the case of sub-optimal DWI. In this study about one-third of MRI examinations were bi-parametric. Theoretically, tumor sensitivity might have been increased if all examinations had been multi-parametric in this heterogeneous MRI cohort, which included some sub-standard DWI protocols.

Image quality is not only dictated by protocol parameters but also to a large extent by patient-related factors such as the ability to avoid motion artifacts in the scanner and to minimize artifacts from intestinal gas and peristalsis. The attention of the MRI technician to these factors may many times determine the difference between good and sub-optimal image quality. Good image quality has to be considered a teamwork between patient and technician. Although the main site MRI was a 10-year-old 1.5 T platform, with standard gradient coils and software, obtained image quality was overall good.

Patient preparation at the main site included butylscopolamine (Buscopan) to reduce intestinal peristalsis. No micro enema was administered. The aim of patient preparations is mainly to reduce artifacts of DWI, but there is a lack of evidence supporting an actual impact on tumor detection [75-78]. Some external sites only used micro enemas, whereas some used only butylscopolamine and some used both.

The lack of fully standardized MRI protocols in this study can also be considered a strength since it gives a picture of how prostate MRI was

performed at the time, and what results could be expected, if the method was widely adopted without harmonized protocols.

Papers II-IV:

The first men were included in G2 trial in 2015. At that time, clinical guidelines recommended prostate MRI mainly as follow-up in men with suspected PC but negative systematic biopsies. The value of MRI to select men for biopsy (the prostate MRI pathway) was not yet established, and prostate MRI was not generally available. With this background, the goal in the G2-trial was to optimize the conditions for a high cancer detection rate in a diagnostic pathway that was still considered experimental. Image exam heterogeneity was minimized by using the same 3T MRI platform (Philips medical systems Achieva dStream) operated by a few MRI technicians, in close collaboration with the study group. The MRI protocol was made compliant to the recently updated PI-RADSv2-guidelines [23], including multi-parametric imaging with DCE and DWI with a high b-value of b$\geq$1400 mm/s^2.

Specific parameters of the G2 MRI protocol are shown in Table 1.

Patient preparations:

Men were instructed to fast for 4 hours and use a self-administered micro-enema 2 h before imaging.

T2W:

Oblique axial T2-weighted images were acquired with a slice thickness of 1.5 mm, exceeding the 3 mm recommendation in PI-RADS. Using a higher than recommended standard is not likely to have significantly aided cancer detection, but adds some flexibility in reconstructing images in arbitrary planes, although is not comparable with reformatting an isotropic voxel acquisition.

The oblique axial plane matching the long axis of the prostate is by some radiologists considered to better depict the prostate anatomy compared to the straight axial plane, matching the long axis of the body. In the G2-trial the oblique axial plane was oriented perpendicular to the bladder floor for historical reasons - to reduce fluid artifacts from the bladder during spectroscopic imaging – even though spectroscopy was never included in the G2-trial protocol. Using the bladder floor as a reference might be sub-optimal since it does not always correlate with the anatomical long axis of the prostate.

Another disadvantage with an oblique plane is that it introduces a risk of variability if serial MRI examinations are performed and imaging planes are not fully matched with previous examinations. However, the same oblique axial plane was used for DWI and DCE to facilitate lesion localization between sequences.

In our current clinical protocol, we have moved to only using straight axial imaging with 3 mm slice thickness for both T2 and DWI.

Coronal and sagittal plane T2-weighted images with a slice thickness of 1.5 mm were acquired orthogonally to the oblique axial plane.

In addition, straight axial T2W images with a slice thickness of 3 mm and a large field of view to cover the entire pelvis from the external iliac arteries to the inguinal regions were acquired to visualize lymph nodes.

DWI/ADC:

The protocol was optimized for high tumor visibility within clinically acceptable scan time and overall optimal contrast-to-noise (CNR) ratio. For DWI (TE, 79 msec) the in-plane resolution was 3 mm and the slice thickness 3 mm resulting in a scan time of 4 minutes. A single-shot echo planar diffusion imaging sequence with two-fold parallel coil acceleration for distortion reduction was used. In the second version of the PIRADS guidelines, a high b-value of b$\geq$1400 sec/mm^2 is mandatory and should preferably be acquired separately [79]. Following the guidelines, four b-values were acquired with 6-fold averaging – b0, b100, b1000, and b1500, three orthogonal directions for each b-value except b0. The ADC map was calculated based on three points, b100, b100 and b1500. Including b1500 is not directly recommended in the guidelines which state that: "The maximum b-value used to calculate ADC is recommended to be $\leq$1,000 sec/mm2 to avoid significant diffusion kurtosis effects that have been described at higher b-values". However, it can be shown in simulations that using a 3-point or more point fit rather than a 2-point fit reduces the b-maximum dependence on ADC. The optimum maximum b-value that results in the highest ADC lesion-to-normal CNR is also increased [74]. Thus, when using the ADC map for visual inspection according to PI-RADS the contrast between tumor and the normal peripheral zone is maintained and no clinical impact on tumor visibility is expected. The high b-value diffusion kurtosis effect (departure from mono-exponential diffusion) affects the measured ADC less, especially if a 3-point fit is used, and is only of clinical importance in quantitative ADC analysis.

Using extrapolated ultra-high b-values ≥2000 has been suggested by several authors as a means to increase tumor conspicuity and reader confidence [80-82]. However, extrapolation or ultra-high b-values were not applied in the G2 trial and is optional according to PI-RADSv2 guidelines.

T1W DCE:

Axial T1W GRE, slice thickness 3 mm was used. Gadolinium-contrast medium (Clariscan, 0.5mmol/mL, GE Healthcare) 0.1 mmol/kg was given intravenously via a power injector at a rate of 3 ml/s. DCE images were acquired for 2.5 minutes with a temporal resolution of 10 s yielding, 15 images per slice, 750 images in total.

Figure 10. 24-sector prostate map. Adapted from the Swedish National Guidelines 2020: "Nationellt Vårdprogram för Prostatacancer" (ref #58)
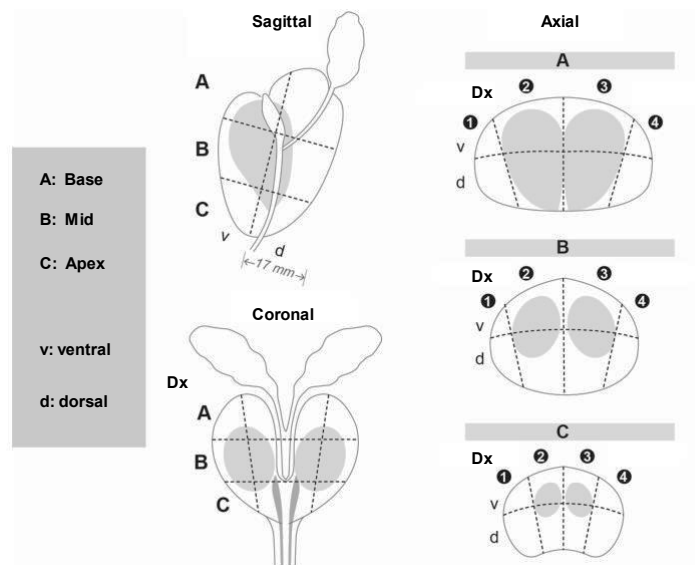

Table 1. MRI protocol G2 study round 1

	T2W axial oblique	T2W coronal	T2W sagittal	T2W axial straight	DWI axial oblique	ADC axial oblique	T1W axial oblique	DCE axial oblique
Pulse sequence	TSE[1]	TSE	TSE	TSE	SPIN	SPIN	TFE[2]	TFE
Comments					b0, b100, b1000 and b1500	b0 excluded	Fat suppressed	Temporal resolution 10s
Scan time (min:sec)	4:16	3:08	3:46	2:51	4:04		0:21	2:30
Slice thickness (mm)	1.5	1.5	1.5	3	3	3	3	3
Space between slices (mm)	1.5	1.5	1.65	3.3	3.3	3.3	3.3	3.3
Rows * columns	432 x 432	512 x 512	512 x 512	384 x 384	92x77 interpolated to 240x240	92x77 interpolated to 240x240	432 x 432	240 x 240
Pixel spacing (mm)	0.42 x 0.42	0.42 x 0.42	0.39 x 0.39	0.52 x 0.52	3 x 3	3 x 3	0.69 x 0.69	1.16 x 1.16
TR (msec)	3906	4275	5141	3570	4000	4130	3.088	3.088
TE (msec)	105	100	100	104	78	79	1.448	1.448

All imaging performed with a 3 T Philips Achieva dStream MRI platform and a pelvic phased array pelvic coil for signal reception. [1]Turbo Spin Echo [2]Dixon FFE (Steady-state Gradient Echo)

2.4 USE OF PROSTATE MRI IN THE STUDIES

PAPER I:

At the hospital where the prostatectomies were performed pre-surgery MRI was incorporated as clinical routine at an early phase, before recommendations in national guidelines were in place, recognizing the potential impact of imaging the prostate but not implementing the imaging information in a structured manner. Thus, it is important to note that included men were selected for prostatectomy based on findings at systematic biopsy and not based on MRI findings. In addition, MRI was not used for detailed radiological assessment of tumor stage or multidisciplinary planning of surgical technique such as degree of nerve-sparing or seminal vesicle invasion.

PAPERS II-IV:

The use of prostate MRI in the G2 trial is radically different from the pre-surgical, "clinical interest"-scenario described in *Paper 1* above. In the G2 trial, men are examined with MRI in case of elevated PSA levels and in two of the study arms biopsies were performed only if MRI showed a lesion to target. The biopsy threshold was originally set to PI-RADS assessment category 3 in any individual pulse sequence but was later changed to overall PI-RADS assessment category 3.

2.5 USE OF PI-RADS AND READING CONSIDERATIONS

For an overview of PI-RADS please see the *Introduction*, chapter 1.2.2.

Paper I:

Several reporting systems were used in the original reports, PI-RADSv1, Likert-scores, and in-house scoring systems. Due to the heterogeneity of the original reports, it was decided to perform a retrospective reading according to PI-RADSv2 which had just been released.

Readers were blinded to previous reports and clinical data.

The study group was familiar with scoring each pulse sequence according to PI-RADSv1. To avoid mistakes in overall scoring - which was a new feature of PI-RADSv2 - it was decided to use the updated PI-RADS syllabus, but to only record the scores of individual pulse sequences, and perform the overall scoring with an automated script.

Up to three lesions per patient were reported by each reader. Each lesion was localized on a 24-sector prostate map as recommended in national guidelines after multi-disciplinary consensus [58]. A weakness of the 24-sector map - compared to the 41-sector map recommended by PI-RADS - is that the sectors are not defined by the zonal anatomy of the prostate, each lesion has to be designated a prostate zone in addition to sector. In Paper 1 the information about the relevant prostate sector was retrospectively registered by reader 1.

As prostate MRI was relatively novel few very experienced readers were available. It was decided to include 2 readers with moderate experience (>200 prior cases), and 1 resident learning prostate MRI (<50 prior cases). The learning curve in prostate MRI may reach an early plateau at around 40 cases, but larger volumes are needed to reach high confidence [83].

Specific methodological limitations:

1. Only a retrospective design was possible since prostatectomy was used as the reference standard.

2. Furthermore, readers were aware that cases were selected due to clinically significant PC and thus were likely to over-report. To reduce bias, a minor proportion of negative cases were added to the reading list, but not included in the statistical analysis. Readers were informed that negative cases had been added but were unaware of the number of negative cases. The proportion of negative cases was still low (10%) compared to clinical pre-biopsy settings, where the negative rate usually is around 50%. Selected negative cases had been scored PI-RADS≤2 by an experienced reader, and were confirmed by benign histology at 12 core systematic biopsies. Using systematic biopsies as the reference standard is a limitation due to systematic under-sampling of ventral regions of the prostate, and the risk of randomly missing smaller tumors in sampled sectors. Thus, saturation biopsies would have been a better reference standard for negative cases.

3. There was a lack of a highly experienced reference reader. It is unclear what level of accuracy could have been reached by an expert reader considering the heterogeneous MRI population.

4. No comparison of results main site versus external sites was made in the published paper. A new analysis was performed for this thesis, please see the *Results* section.

Papers II-IV:

As previously emphasized the main focus at the time of starting up the G2 trial was to maximize tumor detection. Interobserver data was not available for the newly released second version, but it would later be shown that the inter-observer agreement of both PI-RADSv1 and v2 was at best moderate [84-87]. Nevertheless, the study group was concerned that reader variability would limit tumor detection. Thus, it was decided to employ consensus reading defined as reading of each case by at least two of the study radiologists individually before the final reader completed the structured reporting template. In case of different scoring, cases were settled in consensus by at least two of the readers. A clear advantage of consensus reading was that it quickly allowed for the adoption of a common understanding of PI-RADSv2, and for the transfer of knowledge among the 4 radiologists in the group (3 consultants and 1 resident), providing good opportunities to discuss cases. The downsides were increased total reading time, and lack of registered data on the inter-observer agreement and learning curves, since it was considered too time-consuming to let each reader fill in separate protocols, and then transfer all the information to the database manually. Later on in the study a switch from paper protocols to a web-based reporting interface allowed for inter-observer data to be easily

collected, however, this analysis has not yet been performed. In addition, the web-based interface was also built to allow multiple external readers. To date, this option has not been used.

The same 24-sector prostate map described in *Paper 1* was used in *Paper 2-4* to localize lesions, however, the relevant prostate zone was noted on the consensus reporting template (Fig. 10). A maximum of three lesions per patient were recorded. Both individual pulse sequence scores and overall PI-RADS scores were stored in the database for future analysis. The lesion size was measured as the longest diameter, and the diameter perpendicular to the first measurement, usually on axial DWI or T2, as specified by PI-RADS guidelines. Lesion volume was not assessed. Specific measurements of ADC values or pre-defined ADC thresholds were not used. DCE images were assessed visually.

Readers were blinded to PSA levels and assigned study arm.

The study group did not consider alternatives to using PI-RADS. Most international studies adhere to the basic framework of PI-RADS regarding protocols and assessment categories. A different option to the overall scoring in PI-RADS - with decision rules for the dominant sequence and upgrading based on DCE and DWI - is using the Likert score [88]. Grading the likelihood of significant PC with a five-point Likert scale gives the reader a larger degree of freedom compared to the strict PI-RADS algorithm, which sometimes forces the reader to upgrade or downgrade lesions in a questionable way. For instance, as shown in *Paper 3,* the DCE upgrading rules in PI-RADSv2 did not result in increased tumor detection in the screening study. In addition, the Likert score allows the radiologist room to adjust the clinical conclusion according to biochemical data such as PSA levels. In another RADS-system, the LI-RADS (Liver Imaging Reporting and Data System) the final assessment step before finalizing the overall category is "does it seem reasonable?" [89].

In a recent prospective, paired diagnostic study comparing Likert and PI-RADSv2 scoring performed by expert readers, the Likert system resulted in a higher detection rate of clinically significant prostate cancers with a similar number of men biopsied [90]. The study was limited by lack of data on false negatives since men with imaging findings below the biopsy threshold were not biopsied. In the PROMIS study, Likert scoring was used, with a reported sensitivity and specificity of 93% and 41%, respectively, for detection of clinically significant prostate cancer at TPM-biopsies [91].

Using Likert-scoring in the G2-trial could potentially increase the yield of clinically significant prostate cancer given the setting of only highly experienced readers. The G2 trial study design, with an experimental arm including men with a non-standard PSA cut-off (1.8 ng/mL), would however be compromised if PSA levels were unblinded at reading.

Specific methodological limitations:

1. The external validity of the results from our single-center study, including non-standard consensus reading, may not be directly transferable to other centers. A limited external validation including 100 cases distributed over all PI-RADS assessment categories was performed by a single expert reader during the late phase of the first screening round. Cases were retrospectively read and reported using the G2 reporting template (Fig. 12). The observed kappa score was moderate (data to be published).

2. The pretest probability (cancer prevalence) was much lower in the screening cohort compared to published data from clinical cohorts. In a clinical cohort of men under suspicion of prostate cancer the prevalence of clinically significant prostate cancer is commonly 30-40% [54] compared to the low prevalence of <10% reported in a previous pilot study assessing MRI within the tenth round of the Göteborg Randomized Screening Trial (G1) [92].

Assuming that the sensitivity and specificity of the MRI pathway is not significantly affected by the prevalence of cancer, a low prevalence will inherently result in a high negative predictive value (NPV), and a low positive predictive value (PPV). The effects of prevalence on NPV and PPV in a test with high sensitivity (90%) and low specificity (50%), such as prostate MRI, are illustrated in Figure 11.

Consequently, if the radiologist scores MRI as negative in the context of a low-risk population such as screening, there is a high probability of correctly ruling out cancer. On the other hand, the probability/risk of a false positive diagnosis if the radiologist scores MRI as positive is rather high. This is a challenging situation for the radiologist since not missing any cancer is prioritized. But one of the big gains to be expected from using MRI as a prebiopsy filter is to significantly reduce overdiagnosis. The usability of MRI in screening depends on correctly classifying findings as negative or positive with a minimum of indeterminate calls. Indeterminate findings are further discussed in *Paper 3*.

Figure 11. PPV and NPV versus prevalence with test sensitivity of 90% and specificity of 50%.
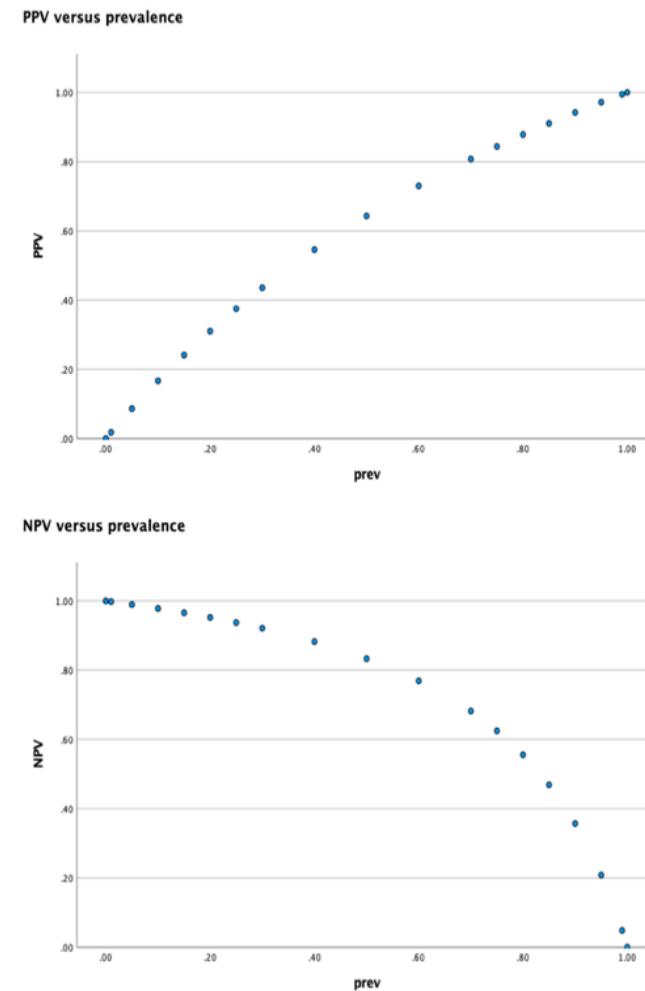
3. We used PRECISE scores to retrospectively assess MRI progression although the cohort was not a true active surveillance cohort as described in PRECISE. However, it was assumed that there were men in the cohort with undiagnosed cancer. We defined significant lesion size progression as > 5 mm in any direction, in accordance with the definition used in the ongoing multicenter active surveillance study, SPCG-17 [93].

Figure 12. G2 trial reporting template including 24-sector prostate map
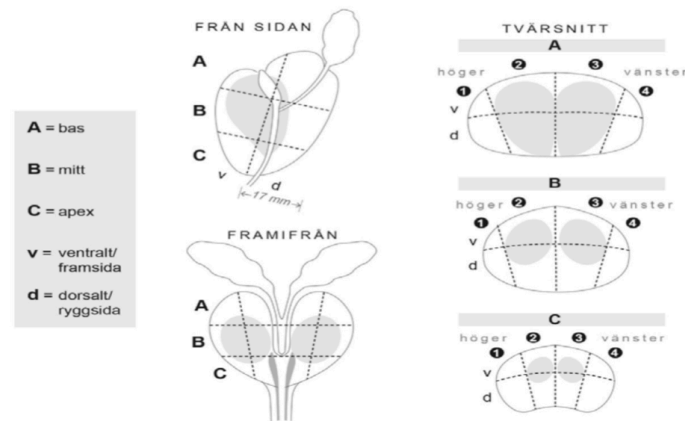
RESULTAT MR-UNDERSÖKNING

Lesion nr: ☐

Personnummer: _____   Namn: _____

Datum undersökning: _____   Granskad av: _____

Finns förändring utanför prostata? Nej ☐   Ja ☐   Om Ja, ange vad: _____

Förändring ses i:
Perifer zon ☐   Transitions-zon ☐   Perifer och transitions-zon ☐

Prostatavolym: _____   Ange förändringarnas lokalisation på skissen:

ESUR-kriterier

*Längd 1 = Maximal längd på misstänkt fynd i mm
*Längd 2 = Vinkelrätt längd mot längd 1 i mm

	T2-weighted MR Imaging	Kryssa i område	Längd 1*	Längd 2*
			Ange misstänkt områdes storlek i mm	
Score 1	Clinically significant disease is highly unlikely to be present			
Score 2	Clinically significant cancer is unlikely to be present			
Score 3	Clinically significant cancer is equivocal			
Score 4	Clinically significant cancer is likely to be present			
Score 5	Clinically significant cancer is highly likely to be present			

	Diffusion weighted MRI	Kryssa i område	Längd 1*	Längd 2*
			Ange misstänkt områdes storlek i mm	
Score 1	Clinically significant disease is highly unlikely to be present			
Score 2	Clinically significant cancer is unlikely to be present			
Score 3	Clinically significant cancer is equivocal			
Score 4	Clinically significant cancer is likely to be present			
Score 5	Clinically significant cancer is highly likely to be present			

Dynamic contrast-enhanced MRI   Positiv ☐   Negativ ☐

Sammanfattande PIRAD-score: ☐

2.6 REFERENCE STANDARD AND DEFINITIONS OF CLINICALLY SIGNIFICANT PC

Paper I:

Definitions at prostatectomy:

Any tumor with **GS 6 and a longest diameter > 10 mm or GS ≥ 7 (any diameter)** was considered ***clinically significant*** (primary definition). Secondary definitions were: GS ≥3+4 (ISUP ≥2) or GS ≥4+3 (ISUP ≥3).

Any tumor **≤ GS 6 and longest diameter < 10 mm** was considered ***clinically insignificant***.

Specific methodological limitations:

1. Using prostatectomy as the reference standard introduces a heavy bias as only men with "surgical" cancer are included. Compared to whole-mount prostatectomy specimens the tumor size is often underestimated at MRI and additional small secondary (<0.5 mL) tumor foci not described at MRI are commonly found. A minor proportion of significant cancers with sparse growth patterns are truly MRI negative [8, 94].

Papers II-IV:

Definition of significant prostate cancer at biopsy:

In the G2 trial, the primary definition of ***significant PC*** is **≥GS 3+4 = 7 (ISUP2)**. A secondary definition is used in *Paper 3* for **GS ≥7 PC excluding small GS3+4 PC with <20% grade 4 and total cancer length <8 mm in ≤4 sectors at systematic biopsies, or tumor volume ≤0.5 ml at prostatectomy.**

Specific methodological limitations:

1. Targeting biopsies at MRI-detected lesions yields more cancer compared to up-front systematic biopsies, but there is also a concern about Gleason inflation driven by the possibility of targeting small foci of higher-grade cancer [40]. In addition, the 2005 ISUP meeting decided that the highest grade should always be reported, regardless of relative percentage [34]. Both these factors

make direct comparisons with prior data based on systematic biopsies more uncertain.

There is currently *no clear consensus* regarding the definition of clinically significant prostate cancer detected in the MRI pathway.

In **PI-RADS** clinically significant PC is defined as *GS ≥ 3+4 = 7 (ISUP 2), and/or volume >0.5 ml and/or extraprostatic extension (EPE).*

In the **PROMIS** study, the primary definition of significant cancer was *GS ≥ 4+3 (ISUP3) or ≥ 6 mm in a single core* and in the **PRECISION** study significant cancer was defined as *GS ≥ 3+4 (ISUP 2)* [49, 91].

2. Systematic biopsies were performed in the reference arm of the G2 trial regardless of MRI findings. However, using systematic biopsies as the reference standard in case of negative MRI is a limitation since the ventral parts of the prostate will not be sampled at all, and other parts will be under-sampled.

Long-term data from the Göteborg 1 screening trial shows that few men with an initial negative systematic biopsy died from PC, but many were diagnosed and treated for PC during the 20 year follow-up [95].

A reference standard with finer granularity than systematic biopsies is saturation biopsies. By using a large number of template cores a small target such as the prostate can be sampled down to every 0.5 cm. In the PROMIS study, 36 core saturation biopsies were used with a reported per-patient sensitivity and specificity of 87% and 47% respectively for detection of ISUP 2 PC. In the same session, 12 systematic biopsy cores were obtained with a reported sensitivity of 60% and specificity of 98% [91]. The PROMIS study was limited by only reporting on a per-patient level, not assessing the spatial correlation of MRI lesions and biopsies. Other studies with transperineal prostate template mapping biopsies (TPM) have reported similar findings compared to the PROMIS study regarding the diagnostic accuracy of MRI to detect clinically significant prostate cancer [96]

To assess tumor extension prior to treatment decision follow-up systematic biopsies were performed in all men with cancer detected at targeted biopsies in Arm 2 and Arm 3 of the G2-trial. In addition, all men in Arm 1 were systematically biopsied according to study protocol as mentioned above.

3. A single pathologist specialized in prostate cancer with 25 years of experience assessed all biopsy cores in the G2 trial. It has to be taken into account that the agreement between pathologists in classifying cases as GS ≥6 versus GS ≤ 7 is not perfect. In a study assessing the agreement of 337 European pathologists in scoring cases as GS ≥ 6 versus GS ≥ 7 with an expert consensus group the mean weighted kappa was only fair/moderate (k=0.41) [97]. At the end of the first screening round 2 external expert pathologists performed a second opinion of all cores with cancer. The full results of this external review are not yet available. In particular, *Paper 3* heavily relies on the distinction between GS=6 PC versus GS ≥7 PC and results may be adjusted according to the external pathology review.

2.6.1 INTERVAL CANCER

Interval cancer may be detected due to clinical symptoms or due to "opportunistic" screening – i.e., additional testing outside of the program in asymptomatic participants. In the G2 Trial interval cancer was defined as any prostate cancer detected in participating men outside of the program regardless of symptoms.

In *Paper IV* men who completed the first screening round with negative MRI and/or negative biopsy, but had prostate cancer diagnosed outside of the program before participating in the second screening round, were considered as diagnosed with interval cancer. Interval cancers were identified by linking with the Regional Cancer Registry.

2.7 BIOPSY TECHNIQUE

Biopsies were performed by a limited number of urologists with 5-25 years of experience with TRUS-biopsies. The biopsy yield of each urologist was continually monitored to assure high quality.

Systematic biopsies were performed transrectally under ultrasound guidance (TRUS) using a standard 12 core sampling scheme (Fig. 13).

Targeted TRUS-biopsies were performed using cognitive fusion. Four cores per lesion were directed at the lesion sector described on the MRI template (Fig. 10 and Fig. 13).

There is no international established consensus regarding the optimal number of targeted cores. The Swedish national guidelines recommend four targeted cores per lesion [58]. In the PRECISION study [49] four cores were targeted per lesion, which is equal to the number of cores used in the G2-trial [72].

Using four targeted cores covering the sector of the suspicious lesion reduces the risk of false-negative biopsies due to targeting error, but does not altogether alleviate the risk. For this reason, men with highly suspicious MRI findings (PI-RADS 5) but negative biopsies in the G2 trial were demonstrated at a multi-disciplinary clinical conference and invited for repeat biopsy after 3 months.

Specific methodological limitations:

The targeting technique used was cognitive, but biopsies were sector-directed rather than truly MRI-directed since the MRI images were not routinely looked at just before the biopsies were performed. Given similar cancer detection rates per PI-RADS assessment category compared to published series, it does not appear that performing sector-directed biopsies is a major limitation.

Software-assisted targeted fusion biopsies were not performed, except in a small number of cases. Randomized studies have not proven superiority of software-assisted fusion over cognitive fusion [45].

Figure 13. Schematic illustration of systematic biopsies versus targeted biopsies. Red circle represents the lesion. Left panel:12 systematic cores cover the dorsal parts of the prostate but misses the lesion in the transition zone on the left anterior side. Right panel: 4 targeted cores directed at the lesion.
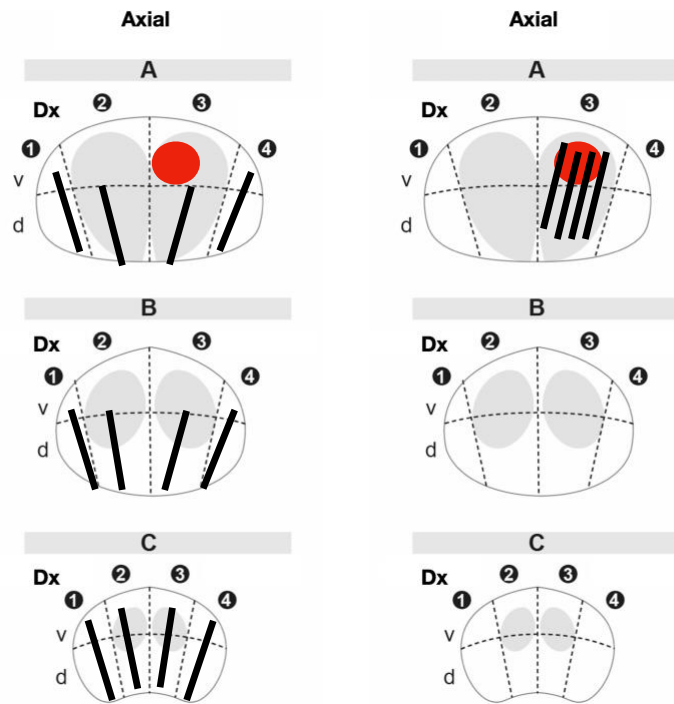
2.8 STATISTICAL METHODS

***Paper I*: Assessing PI-RADS interobserver agreement with the Kappa statistic**

In radiology interobserver agreement is usually reported with the Kappa statistic (κ) [98]. The κ-score adjusts the percentage of agreement between two readers in categorizing events - for instance cancer yes/no - by adjusting for the probability that the agreement occurred by chance.

$$\kappa = (\text{observed agreement} - p_{chance})/(1\text{-}p_{chance})$$

The probability of agreement by chance is calculated by adding the probabilities of the observers answering yes/no:

$$p_{chance} = (\% \text{ Observer A } yes * \% \text{ Observer B } yes) + (\% \text{ Observer A } no * \% \text{ Observer B } no)$$

The level of agreement with κ-scores was assessed as follows: slight agreement 0.01-0.20; fair agreement 0.21-0.40; moderate agreement 0.41-0.60; substantial agreement 0.61-0.80 and almost perfect agreement 0.81-1.

An effect of the Kappa statistic is that the same percentage of agreement can produce different κ-scores if agreement is mostly in one class compared to if agreement is evenly distributed among classes [99]. In other words, the probability of agreement by chance is influenced by the distribution between classes. By also reporting the actual percentage of agreement the level of chance-adjustment in the κ-scores becomes more evident.

The weighted κ-score is used to add weighting to the level of disagreement. For example, in the ordinal 5-point PI-RADS assessment scale a higher weighting could be given to a 2 versus 4 disagreement, compared to a 3 versus 4 disagreement. Weighted κ-scores were not used in this paper, instead PI-RADS scores were dichotomized with a threshold at PI-RADS 3 or PI-RADS 4.

The baseline κ assumes that observations are randomly allocated. Thus, the κ-scores is a ratio that "hides allocation". If allocation is clinically relevant an alternative is to use FROC-analysis [100]. In prostate MRI agreement on the localization of a lesion is important in for example in biopsy planning. In a

future interobserver study FROC-analysis could be an interesting methodological approach.

Paper II: **Testing for non-inferiority - bpMRI versus mpMRI**

Non-inferiority testing is used to demonstrate that a method is non-inferior compared to a reference method. This requires a somewhat different analytical approach compared to testing for superiority. If the lower bound of the confidence interval is above the pre-specified non-inferiority limit, non-inferiority is claimed (Fig. 14) [101]. The non-inferiority margin should be justified clinically and statistically.

Figure 14. Illustration of non-inferiority testing: Point estimate of the relative risk
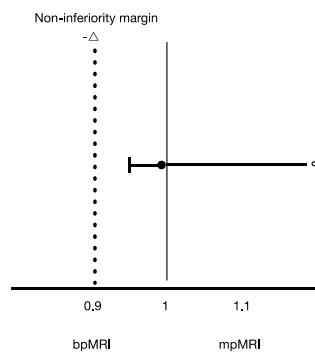
(black dot) close to 1 and lower bound of confidence interval above non-inferiority margin (dotted line). One-sided test.

In our case, we wanted to show that bpMRI was non-inferior to mpMRI. If their detection rates were equal the relative risk had been 1. Our chosen 10% non-inferiority limit was greater than the 5% limit recommended by the START recommendations when reporting non-inferiority in detection of significant PC [47]. However, in this study, we considered the 10% limit as clinically relevant as we tested a low-risk screening population for any cancer with a Gleason score ≥6.

We used a 5% *significance level* which means that there is a 5% risk of falsely reporting bpMRI non-inferior to mpMRI for the given non-inferiority margin (type 1 error). This could be interpreted as 95% confidence that bpMRI will detect at least 90% of lesions detected with mpMRI.

A *power* of 80% means that there is a 20% risk of not being able to reject that bpMRI is not non-inferior to mpMRI, when it actually is non-inferior (type 2 error). The power calculation in Paper 2 was based on "paired binomial data" which means that not only the individual detection rates of bpMRI and mpMRI were important, but also the proportion of common assessments. The assumption of 97% common assessments was based on a previous study by Kuhl reporting 99% common assessments in scoring bpMRI and mpMRI [17]. The true positive rate of 15% (number of men with positive targeted biopsy/total number of targeted biopsies) was estimated from previous data from the G2 trial.

The sample size needed for a full randomized trial was calculated to approximately 12600 MRI examinations to prove bpMRI non-inferiority to mpMRI, even with a 10% non-inferiority limit (15% prevalence, 80% power, 5% significance level).

By performing both tests in the same man (paired diagnostic study design) the required sample was reduced to 550, with the same prerequisites.

The fact that the actual observed point estimate and lower bound of the one-sided 95% CI was clearly above the test limit (-10%), adds further weight to the claim of bpMRI non-inferiority.

Figure 15. Example of a Receiver Operating Characteristic (ROC) curve. PI-RADS scores versus detection of PC in first round of the G2 trial. AUC = 0.74.
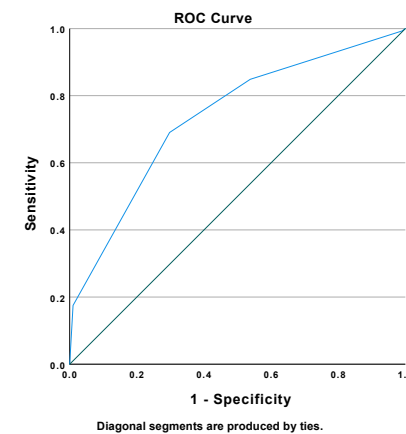
Paper III. Multivariable logistic regression models to assess lesion size, DCE, and PSAD as predictors of cancer

Logistic regression is the "default" method for multivariable modeling if the outcome is binary, such as cancer yes/no. If the outcome is a continuous variable, for example size, _linear regression_ is the commonly used method. As a rule of thumb, approximately 10 events are needed per variable in a multivariable regression model. The event is defined as the least common outcome. To exemplify: In our case, the event was "detected cancer" but in a different theoretical study population with 90% cancer prevalence the event would have been "no cancer detected".

If the distribution of observations is uneven, as often is the case in biology with many low values and a few high values (right-skewed distribution), the few high values can cause disproportionate effects in the model. A possible solution is to log transform the data in order to even out the distribution. We log transformed variables PSAD and lesion area with base 2, but any log base could have been used.

The _Receiver Operating Characteristics_ curve (ROC) is a graph with true positive rate on the y-axis and false positive rate on the x-axis (Fig. 15). The graph visualizes the performance of a classification model at all thresholds. The _area under the curve_ (AUC) ranges from 0 to 1. If the AUC is equal to 0.5 the model is useless for classification, and if the AUC is equal to 1 classification is perfect. The AUC is both scale-invariant (only a measure of how well the model ranks events) and threshold-invariant (includes all thresholds) allowing for easy comparison of models.

If the event rate is low overfitting of the data to the model can be problematic, especially if the models are used to recommend specific thresholds, for example, to find a reliable PSA-density cut off to select men for biopsy. One way of assessing overfit in a (logistic) regression model is to adjust AUC for optimism [102]. A small overfit of data is generally not problematic if the model is used only for assessing the potential usability of a variable.

Papers I-IV. P-values and confidence intervals

The _p-value_ is the probability of obtaining a test result at least as extreme as the observed if the null hypothesis is true. In a "standard", two-sided test the null hypothesis is that the groups are equal. In a non-inferiority (one-sided) test the null hypothesis is that the tested method is equal to the pre-specified margin, or worse.

To exemplify: If a p-value of 0.05 is considered statistically significant, obtaining a p-value above 0.05 - for instance - 0.15, means that the null hypothesis cannot be rejected. On the other hand, if the p-value was below 0.05, the null hypothesis is rejected but it does not mean that there is a 95% likelihood that the alternate hypothesis is actually true, only that the observed result would be highly unlikely if the null hypothesis was true [103].

Confidence intervals contain more information than p-values as they supply information about the precision of the point estimate by estimating a range of values containing the true (but unknown) parameter value 95 out of 100 times. There are most often different theoretical methods for calculating confidence intervals based on known distributions. Boot-strapping is an alternative approach that can be used if a valid method (distribution) is lacking. In boot-strapping, the observed data is resampled many times in order to estimate the distribution.

3 RESULTS

PAPER I:

The study cohort consisted of 97 men who were examined with MRI before prostatectomy.

With a threshold of PI-RADS ≥3, the index tumor detection rates for the three readers were 76%, 67% and 76% respectively. Considering ISUP ≥2 PC the detection rates were 76%, 73%, and 80 % for the three readers respectively (*Table 2 in the published manuscript*).

With a threshold of PI-RADS ≥4 the index tumor detection rates were lower, 63%, 54%, and 66% respectively (*Table 3 in the published manuscript*). Considering ISUP≥2 PC the detection rates were 64%, 64%, and 74% for the three readers respectively (*Table 3 in the published manuscript*).

The interobserver agreement between pairs of readers was fair to moderate considering lesions scored PI-RADS ≥3, with kappa coefficients of 0.38 (95% CI 0.18-0.58), 0.31 (95% CI 0.09-0.53), and 0.54 (95% CI 0.35-0.72) (Table 4 in the published manuscript). The corresponding percent agreements were 87.1%, 83.1%, and 81.7% for a PI-RADS score of ≥3 and 48.8%, 47.8% and 78.3% for a PI-RADS score of ≤2 (*Table 2, previously unpublished results*).

The interobserver agreement between pairs of readers was moderate considering lesions scored PI-RADS ≥4, with kappa coefficients of 0.50 (95% CI 0.32-0.67), 0.49 (95% CI 0.31-0.68), and 0.54, respectively (95% CI 0.37-0.71) (*Table 4 in the published manuscript*).

Table 2.

	Reader 3 PI-RADS ≤2	Reader 3 PI-RADS ≥3	Total
Reader 1 PI-RADS ≤2	11 (47.8%)	12 (16.9%)	23
Reader 1 PI-RADS ≥3	12 (52.2%)	59 (83.1%)	71
Total	23	71	94*

A) Percent agreement Reader 1 vs Reader for 3 PI-RADS≥3 (Kappa score 0.31; 95% CI 0.09-0.53)

	Reader 2 PI-RADS ≤2	Reader 2 PI-RADS ≥3	Total
Reader 1 PI-RADS ≤2	15 (48.8%)	8 (12.9%)	23
Reader 1 PI-RADS ≥3	16 (51.6%)	54 (87.1%)	70
Total	31	62	93**

B) Percent agreement Reader 1 vs Reader 2 for PI-RADS≥3 (Kappa score 0.38; 95% CI 0.18-0.58)

	Reader3 PI-RADS ≤2	Reader 3 PI-RADS ≥3	Total
Reader 2 PI-RADS ≤2	18 (78.3%)	13 (18.3%)	31
Reader 2 PI-RADS ≥3	5 (21.7%)	58 (81.7%)	67
Total	23	71	94***

C) Percent agreement Reader 2 vs Reader 3 for PI-RADS≥3 (Kappa score 0.54; 95% CI 0.35-0.72)

*3 cases missing
** 4 cases missing
*** 3 cases missing

Table 3 shows a similar distribution of PI-RADS-scores when comparing MRI performed at the main sites and MRI performed at external sites, 12% versus 10% negative MRI examinations and 88% versus 90% positive MRI examinations. Results not previously published.

Table 4 shows similar distributions of ISUP-scores (at prostatectomy) between men who had their MRI at the main site compared to external sites, 23% versus 24% for ISUP 1, 50% versus 48% for ISUP 3, and 26% versus 29% for ISUP ≥3. Results not previously published.

Table 3. Dichotomized distribution of PI-RADS-scores at main site versus external sites.

	All readers PI-RADS ≤2	At least one reader PI-RADS ≥3	Total
Main site	8 (12%)	58 (88%)	66
External sites	3 (10%)	28 (90%)	31
Total	11	86	97

Table 4. ISUP scores at prostatectomy - main site versus external sites.

	ISUP 1	ISUP 2	ISUP ≥3	Total
Main site	16 (24%)	33 (50%)	17 (26%)	66
External sites	7 (23%)	15 (48%)	9 (29%)	31
Total	23	48	26	97

PAPER II.

The final study cohort included 551 men who were examined with prebiopsy mpMRI due to screening-detected increased PSA levels *(Table 1 of the published manuscript)*.

The distribution of PI-RADS scores is shown in *Tables 2 and 3* of the published manuscript. In summary, bpMRI was negative (PI-RADS ≤2) in 423 out of 551 (77%) and positive in 128 out of 551 (23%) of cases respectively and mpMRI was negative (PI-RADS ≤2) in 415 out of 551 (75%) and positive in 136 out of 551 (25%) of cases respectively.

With mpMRI, the frequency of detected cancer was 84/551 (15.2%; 95% CI: 12.4-18.4) and with bpMRI, the frequency of detected cancer was 83/551 (15.1%; 95% CI: 12.3-18.2). The relative risk of detecting cancer comparing bpMRI with mpMRI was 0.99 (95% one-sided CI: >0.95). Hence, the lower bound of the one-sided CI was above the pre-specified limit of 90% (non-inferiority margin of 10%) and bpMRI was proven non-inferior *(Table 5 in the published manuscript)*.

The number of false positives was 45/128 (35.2%) with bpMRI and 52/136 (38.2%) corresponding to a relative risk of 0.92 (95% CI: 0.84-0.98). Out of the 8 lesions only scored positive with mpMRI 7 lesions were false positives.

The positive predictive value for detecting cancer with MRI followed by targeted biopsies was 83/128 (64.8%) with bpMRI and 84/136 (61.8%) with mpMRI; relative risk = 1.05 (95% CI: 1.01-1.10).

PAPER III.

In the first screening round, a total of 2150 mpMRI examinations were performed up until 20 October 2020. The distribution of PI-RADS scores versus Gleason scores prior to external pathology review is shown in *Table 5*. Results not previously published.

Table 5.

	No PC	3+3	3+4	4+3	3+5	4+4	4+5	5+4	Total
PI-RADS ≤2	1438	53	10	1	0	0	0	0	1502
PI-RADS 3	134	47	14	1	0	1	1	0	198
PI-RADS 4	152	130	65	15	1	4	2	0	369
PI-RADS 5	7	13	40	10	0	8	1	2	81
Total	1731	243	129	27	1	13	4	2	2150

Distribution of PI-RADS scores and Gleason scores in the first screening round up to 20 October 2020.

The total number of men with peripheral zone index lesions scored PI-RADS 3-4 was 502. A subset of 280 men scored DWI 3 and out of these 253 were biopsied up to 20 October 2020 *(Fig. 1 in the manuscript)*.

Out of 88 PC detected in total, 85 were detected with targeted biopsies and 3 were detected only with systematic biopsies in a sector not matching the same or adjacent sector of the MRI findings.

Based on the final pathology results, after the external review 58/253 (23%) were ISUP1 PC and 30/253 (12%) were ISUP ≥2 PC.

Out of 253 lesions, 132 (52%) were DCE positive; 79 with benign histology, 34 with ISUP1 PC, and 19 with ISUP ≥2 PC. In the case of positive DCE (DCE upgrading) the positive predictive value was 40% (95% CI: 32-49) and 14% (95% CI: 9-22) for any ISUP PC and ISUP ≥2 PC respectively.

In *Tables 2a and 2b* of the manuscript, the AUC and LR test results of multivariable regression models are shown.

No statistically significant improvement of the prediction of any ISUP PC or ISUP≥2 PC was obtained by models including a) PSAD and lesion area; b) PSAD and DCE; c) PSAD, lesion area, and DCE compared to a model including only PSAD (Tables 2a and b in the manuscript). PSAD alone was statistically significantly correlated to both any ISUP PC and ISUP≥2 PC (p<0.001). AUC:s for detection of any ISUP PC and ISUP≥2 PC respectively with predictors alone were: PSAD 0.66 and 0.73; lesion area 0.54 and 0.57; DCE 0.56 and 0.56.

In *Figure 16* the proportion of detected PC is plotted against a range of PSAD cut-offs and the corresponding proportion of men biopsied, to illustrate the potential clinical value of PSAD. Using a 0.1 ng/ml^2 PSAD cut-off would result in detecting 40% of any ISUP PC, and 57% of ISUP ≥2 PC with a 75% reduction of biopsies in our cohort. However, the low event rate of ISUP ≥2 PC prevents establishing a clinically reliable cut-off.

A secondary definition of significant PC excluding low volume GS 3+4 PC is described in the *Methods* section. With this definition, the number of significant PC was 20/253 (8%). A PSAD cut-off of 0.1 ng/ml^2 would result in the detection of 70% of these higher-risk cancers with a 75% reduction of men biopsied *(Fig. 2 in the manuscript)*.

Figure 16. Proportion of men with PC detected versus PSA-density and proportion of men biopsied. Red and black dotted lines shows that in our study, a 0.1 ng/ml² PSAD cut off would result in detecting 57% of GS≥7 cancers and 40% of any GS cancers with biopsy of 25% of men.
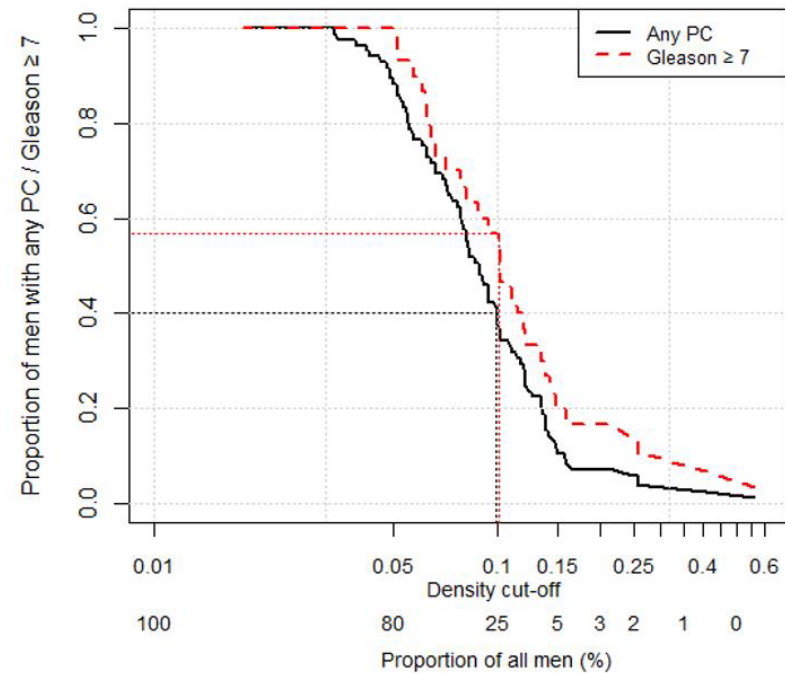
PAPER IV:

Up to 30 June 2020, a total of 474 men had completed both the first round of screening - with negative MRI or positive MRI with negative biopsies - and participated in the second round of PSA-screening after 2 years.

In total 78 out of 474 men were biopsied in the second round, resulting in the detection of 23 PC (5%) of any ISUP grade including 5 ISUP ≥2 PC (1%). Further patient characteristics are shown in *Table 1* of the manuscript.

The majority of men (81%) had elevated PSA levels also at follow-up, and were scheduled for MRI. Out of these 383 men 376 completed MRI with diagnostic results. The final MRI cohort consisted of 376 men *(Fig. 1 of the manuscript)*.

Considering PI-RADS ≥3 as positive MRI a total of 104 men had a positive MRI at least once in the first, or second rounds, and these MRI examinations were re-read according to the PRECISE guidelines. Out of these 104 men, 19 were only MRI positive in the second round, and 6 were diagnosed with PC. Another 22 men were MRI positive in both rounds, and 8 of these men were diagnosed with PC *(Fig. 2 of the manuscript)*.

Results of the re-read are shown in *Table 2* of the manuscript. A majority of intervals (71%) were stable, i.e., PRECISE 3. Regression occurred in 20% of intervals, i.e., PRECISE 1-2 and progression occurred in 8% of intervals, i.e., PRECISE 4. Second round targeted biopsies (prior to PRECISE-scoring) yielded 14 cancers in the group of 104 men, 12 PC in PRECISE 3, and 2 PC in PRECISE 4. In total 5 out 12 PC were ISUP 2-3, all PRECISE 3.

In the above-mentioned group with positive MRI only in the second-round, 3 out of 19 men were diagnosed with ISUP 2 PC at second-round biopsy, but retrospectively considered as first-round MRI reading failures (missed lesions) with a stable appearance at follow up - PRECISE 3.

In the group with positive MRI at both baseline and follow-up, 2 out of 22 men were diagnosed with ISUP 2-3 PC at second-round biopsy, both PRECISE 3 and considered as first-round biopsy failures.

Baseline PI-RADS scores from the re-read versus PRECISE scores and number of detected ISUP ≥2 cancers at second-round biopsies are summarized in *Table 6*.

Table 6. PI-RADS overall score at baseline MRI (re-read) versus PRECISE-scores.

	PRECISE ≤2	PRECISE 3	PRECISE ≥4	Total
PI-RADS ≤2	5	47 **2 PC ISUP 2**	7	59
PI-RADS ≥3	17	27 **3 PC ISUP 2-3**	1	45
Total	22	74 **5 PC ISUP 2-3**	8	104 **5 PC ISUP 2-3**

In bold face number of ISUP ≥2 cancers diagnosed in the second round.

Interval cancer, i.e., cancer detected outside the regular screening program, was diagnosed in 10 men (2 ISUP grade 5, 1 ISUP grade 2, and 7 ISUP grade 1). Both high-risk cancers were detected with MRI (PI-RADS 4 and PI-RADS 5) but missed at initial biopsy *(Table 3b of the manuscript)*.

4 DISCUSSION

Strategies for optimizing the use of MRI in a PSA-based screening program for prostate cancer can be broadly summarized in two categories – optimizations of *protocols* and optimizations of *diagnostic pathways*.

The *MRI protocol* should have high tolerability for the participants and not be too time-consuming, invasive, or costly.

The *diagnostic pathway* should take into account that MRI acts as a pre-biopsy filter in a low-risk population with high expectations thresholds, both not to cause harm by overdiagnosis, and to accurately detect clinically significant cancer. A good safety net in case of failures of MRI or biopsies should be built into the system with well-calibrated re-testing intervals.

Ad Paper I. Interobserver agreement and tumor detection in pre-operative MRI with non-expert readers

Key results: The average index tumor detection rate was 70% and the average inter-observer agreement was fair to moderate with a κ-score of 0.41.

An often-discussed limitation of PI-RADSv2 is the moderate interobserver agreement in scoring, despite efforts to standardize reading and reporting. Our results are not substantially different from other studies including both expert and non-expert readers. In a study including six experienced readers from academic centers, the overall agreement was moderate (κ-score 0.46) [104]. In another study including nine radiologists with different levels of experience, the overall agreement was moderate for detection of index lesions (κ-score 0.42) [105].

In PI-RADSv2.1 further adjustments were made to increase reader agreement but in a study with six radiologists of different experience levels the overall agreement was still only at the low end of moderate (κ-score 0.42) [106].

The index tumor detection rate was overall slightly lower compared to the per lesion sensitivity of 76-84% reported in three previous studies with prostatectomy as the reference standard [24, 94, 107]. The considerable heterogeneity of included MRI examinations and non-expert reader setting might have contributed to a lower detection rate in our study.

Implications for screening: Quality controls are of paramount importance to assure adherence to recommended imaging protocols, and to assess reader

performance. Given the large reader variability for both experts and non-experts in this and other studies, it seems pertinent to encourage readers to attain a high level of experience, but also to, on a regular basis, calibrate PI-RADS-scoring within the reading group as well as externally with other groups.

Ad Paper II. Evaluation of bi-parametric MRI protocol in screening

Key results: Cancer detection with bpMRI was non-inferior to mpMRI in screening with PSA followed by MRI.

During the last couple of years, there has been an ever-increasing interest in bi-parametric MRI, with several systematic reviews and meta-analyses reporting similar cancer detection rates with both methods [108-111]. However, the number of blinded head-to-head studies is limited, and there is presently no data available from large-scale multicenter studies [112]. Our results heavily support the use of bpMRI in screening based on our prospective, blinded study design and well-defined screening cohort.

Implications for screening: Bi-parametric MRI should be the method of choice in screening. Bi-parametric MRI is non-invasive without exposure to gadolinium contrast and reduces both room-turn-over time and cost.

Ad Paper III. Evaluation of contrast enhancement, lesion size, and PSAD in screening-detected indeterminate peripheral zone lesions (PI-RADS 3).

Key results: More than 40% of all MRI examinations scored as positive (PI-RADS ≥3) in the first-round screening cohort had peripheral zone PI-RADS DWI =3 index lesions, but the relative yield of prostate cancer was low (35% had any PC and 12% had PC ≥ISUP 2). Cancer detection was statistically significantly correlated to PSAD, but not to contrast enhancement or lesions size.

Most comparable studies do not separately report the frequency and tumor detection rate in peripheral zone PI-RADS 3 lesions with a DWI-score of 3. However, our results are close to previously reported frequencies and cancer detection rates for the entire PI-RADS 3 group. In a review including 7 studies with biopsy naïve men the percentage of PI-RADS 3 lesions ranged from 14-39 %, and the yield of significant PC ranged from 4-27% [113]. In the PRECISION trial the frequency of PI-RADS 3 was 20%, and the detection rate of any PC and PC ISUP ≥2 was 34% and 12% respectively [49].

We did not validate the use of contrast enhancement to predict prostate cancer in indeterminate lesions (DCE-upgrading). Upgrading based on DCE has only been strongly validated in studies with prostatectomy as the reference standard [114, 115], but studies with biopsies as reference standard have shown varying results [18, 30, 116].

Generally, a PSA density of <0.1 ng/ml^2 is correlated with a low risk of prostate cancer, and a density >0.2 ng/ml^2 is correlated with a high-risk [117-120]. Previous MRI studies have shown that a PSAD <0.15 ng/ml^2 correlates with a low risk of prostate cancer in men with PI-RADS 3 [121-123]. In our study, a lower than standard theoretical PSAD threshold of 0.1 ng/ml^2 would result in reducing the number of biopsies by 75% while detecting 57% of ISUP≥2 PC, and 70% of ISUP≥2 PC excluding small cancers with <20% grade 4 and no grade 5.

Implications for screening: Two out of five men subjected to targeted biopsies in the G2 study had peripheral zone DWI=3 lesions. Establishing a reliable screening-specific PSAD threshold is a top priority considering the potential gains from reducing biopsies and risk of overdiagnosis. It could also be argued that a screening-specific PSAD threshold should take into account that delaying diagnosis of small intermediate-risk cancers may even be beneficial, provided that these cancers will be detected in a curable stage at follow-up screens. Detection of the few larger size significant cancers nested in the group must however not be delayed.

Ad Paper IV. Two-year MRI follow up of men not diagnosed with PC in the first round

Key results: Both benign and malignant appearing lesions are mostly stable at 2 year follow-up. In men with negative first-round MRI very few cancers were diagnosed in the second round. The incidence of high-risk interval cancer was low, and these cancers were MRI positive but biopsy negative.

There are only a few previous studies describing the natural history of MRI lesions with few worrisome features. In one study including 153 men with small index lesions (≤7 mm) and benign histology, or low-risk PC at biopsy, lesions did not significantly increase in size over a 2 year follow-up period [124]. In another study including 83 men with serial MRI, either for diagnostic purposes or during active surveillance, changes in PI-RADS score and/or lesion size were small during a mean follow-up of 1.9 years [125].

The frequency of MRI progression in active surveillance (AS) is substantially higher. In a study including 553 men with low or intermediate-risk PC managed by AS the frequency of MRI progression, defined as PRECISE score ≥4 was 43%. Out of men with MRI progression 61% also had clinical progression, defined as progression to ISUP ≥3 or initiation of active treatment [126]. However, in a different study the risk of clinical progression was low in men with low-risk PC (ISUP1) and stable imaging (PRECISE≤3) [127].

Implications for screening: The 2-year interval in case of negative MRI appears safe and is potentially extendable to 4 years. In men with worrisome, and highly worrisome MRI findings (PI-RADS 4-5) a closer follow-up is warranted, as a few high-risk interval cancers were missed at first-round biopsies. To reduce the risk of diagnostic failures, highly suspicious MRI lesions with negative biopsies should be demonstrated at rounds and scheduled for early re-biopsy.

5 FUTURE PERSPECTIVES

The test characteristics of prostate MRI are ideal for screening with a very high sensitivity to detect clinically significant prostate cancer and inherently high negative predictive value if used in a low prevalence setting. Still, there is ample room for improvements. Many cancers diagnosed in the first screening round were of low risk and a considerable proportion of imaging findings were indeterminate.

Several strategies for improving the future use of MRI in screening could be suggested.

Considering the large reader variability and difficulties in predicting the desired binary outcome of significant cancer or no cancer improvements should focus on developing quantitative, reproducible MRI parameters stable across platforms. For instance, continued work on the use of ADC ratios and fully quantitative cross-platform stable ADC measurements could increase reader confidence and diagnostic yield as well as set the stage for future artificial intelligence (AI) developments [128].

With the ever-increasing case-load AI will surely find a role in prostate MRI, especially in screening. Considering that almost 60% of MRI examinations in the first round of the G2 trial were scored as negative an AI solution to identify negative cases would be extremely helpful. Automatization of tedious tasks such as assessing prostate volume and lesion size/volume in serial examinations should lie right around the corner. Further down the road applications for lesions detection/double reading as well as detection of significant incidental findings outside the prostate could be of help to the radiologist in clinical practice.

Another major challenge with screening is a timely diagnosis, i.e., finding cancer at a curable stage but not too early considering the risk of reduced quality of life after active treatment. Thus, the risk of being over-diagnosed and overtreated is not revoked with MRI. Both surgery and irradiation can lead to long-term complications such as erectile dysfunction and/or incontinence, a high price to pay for the individual man if the treatment was not warranted in the first place. The psychological burden of a cancer diagnosis must also be considered if the cancer is managed by active surveillance.

Future work should focus on developing algorithms for predicting tumor extent and risk of adverse pathological outcomes in the case of positive MRI [129]. With better prediction models biopsies might be deferred in men with imaging

features suggestive of low-risk cancer and surveillance could be performed with only imaging until features suggestive of more aggressive cancer occur.

The G2 Trial provides important data on the effects of combining PSA testing with MRI followed by targeted biopsies in a screening program for prostate cancer. We expect a significant reduction of overdiagnosis of indolent prostate cancer compared to PSA followed by only systematic biopsies. However, it will take at least another 5 years before the first results on mortality in the screening group compared to the control group can be analyzed and another decade or two before the full effects can be evaluated.

The final answer to the big screening question still lies in the future. With the ever-expanding realm of technical possibilities at hand, the thought that we could cure cancer before it produces symptoms tickles the mind, but the risk of harming instead of helping is not to be taken lightly. The best path up the mountain can only be found by performing well-designed randomized trials. It is extremely rewarding to be part of such an effort.

6 CONCLUSIONS

1: Non-expert readers detected a majority but not all prostate cancers with preoperative MRI. The agreement between readers was fair to moderate highlighting the importance of a quality assurance program.

2: Cancer detection with bi-parametric MRI was non-inferior to multi-parametric MRI in a PSA-based screening program for prostate cancer.

3: In screening, bi-parametric MRI should be the method of choice as it also reduces room turn-over time and saves healthy men exposure of gadolinium contrast agents.

4: In screening, the frequency of significant prostate cancer in indeterminate peripheral zone MRI lesions (PI-RADS 3) was low and unrelated to contrast enhancement and lesion size.

5: Selecting men with screening-detected indeterminate peripheral zone MRI lesions for biopsy based on PSA-density is a promising approach but more data is needed to establish a reliable cut-off.

6. Very few men with a negative first-round MRI were diagnosed with clinically significant prostate cancer in the second round indicating that a 2-year interval is safe and it further suggests that an extended interval could be considered in this group.

7. The majority of MRI lesions were of stable imaging appearance at two year follow-up. A closer follow-up of men with PI-RADS 4 and 5 lesions with negative biopsies is, however, warranted as the risk of significant prostate cancer in this group of men was not negligible.

REFERENCES

1. Welch, G., ed. *Overdiagnosed. Making people sick in the pursuit of health.* 2011.
2. Steyn, J.H. and F.W. Smith, *Nuclear magnetic resonance imaging of the prostate.* Br J Urol, 1982. **54**(6): p. 726-8.
3. Hricak, H., et al., *MR imaging of the prostate gland: normal anatomy.* AJR Am J Roentgenol, 1987. **148**(1): p. 51-8.
4. Brown, G., et al., *The role of intravenous contrast enhancement in magnetic resonance imaging of prostatic carcinoma.* Clinical Radiology, 1995. **50**(9): p. 601-606.
5. Giganti, F., et al., *The Evolution of MRI of the Prostate: The Past, the Present, and the Future.* AJR Am J Roentgenol, 2019. **213**(2): p. 384-396.
6. Issa, B., *In vivo measurement of the apparent diffusion coefficient in normal and malignant prostatic tissues using echo-planar imaging.* J Magn Reson Imaging, 2002. **16**(2): p. 196-200.
7. Sillerud L.O, H.K.R., Griffey R.H, *Spectroscopy of the human prostate* MRIM, 1988. **8/1988**: p. 224-230.
8. Langer, D.L., et al., *Prostate tissue composition and MR measurements: investigating the relationships between ADC, T2, K(trans), v(e), and corresponding histologic features.* Radiology, 2010. **255**(2): p. 485-94.
9. De Visschere, P.J., et al., *Multiparametric magnetic resonance imaging characteristics of normal, benign and malignant conditions in the prostate.* Eur Radiol, 2017. **27**(5): p. 2095-2109.
10. Franiel, T., B. Hamm, and H. Hricak, *Dynamic contrast-enhanced magnetic resonance imaging and pharmacokinetic models in prostate cancer.* Eur Radiol, 2011. **21**(3): p. 616-26.
11. Tan, C.H., et al., *Dynamic contrast-enhanced MRI for the detection of prostate cancer: meta-analysis.* AJR Am J Roentgenol, 2015. **204**(4): p. W439-48.
12. Rosenkrantz, A.B. and S.S. Taneja, *Radiologist, be aware: ten pitfalls that confound the interpretation of multiparametric prostate MRI.* AJR Am J Roentgenol, 2014. **202**(1): p. 109-20.
13. McDonald, R.J., et al., *Gadolinium Retention: A Research Roadmap from the 2018 NIH/ACR/RSNA Workshop on Gadolinium Chelates.* Radiology, 2018. **289**(2): p. 517-534.
14. Turkbey, B., et al., *Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2.* Eur Urol, 2019. **76**(3): p. 340-351.

15. Thestrup, K.C., et al., *Biparametric versus multiparametric MRI in the diagnosis of prostate cancer.* Acta Radiol Open, 2016. **5**(8): p. 1-8.

16. Stanzione, A., et al., *Biparametric 3T Magnetic Resonance Imaging for prostatic cancer detection in a biopsy-naive patient population: a further improvement of PI-RADS v2?* Eur J Radiol, 2016. **85**(12): p. 2269-2274.

17. Kuhl, C.K., et al., *Abbreviated Biparametric Prostate MR Imaging in Men with Elevated Prostate-specific Antigen.* Radiology, 2017. **285**(2): p. 493-505.

18. Choi, M.H., et al., *Prebiopsy Biparametric MRI for Clinically Significant Prostate Cancer Detection With PI-RADS Version 2: A Multicenter Study.* AJR Am J Roentgenol, 2019. **212**(4): p. 839-846.

19. Kim, Y.J., J.S. Huh, and K.K. Park, *Effectiveness of Bi-Parametric MR/US Fusion Biopsy for Detecting Clinically Significant Prostate Cancer in Prostate Biopsy Naive Men.* Yonsei Med J, 2019. **60**(4): p. 346-351.

20. Zawaideh, J.P., et al., *Diagnostic accuracy of biparametric versus multiparametric prostate MRI: assessment of contrast benefit in clinical practice.* Eur Radiol, 2020. **30**(7): p. 4039-4049.

21. Wallström, J., et al., *Bi- or multiparametric MRI in a sequential screening program for prostate cancer with PSA followed by MRI? Results from the Göteborg prostate cancer screening 2 trial.* European Radiology, 2021.

22. Barentsz, J.O., et al., *ESUR prostate MR guidelines 2012.* Eur Radiol, 2012. **22**(4): p. 746-57.

23. Barentsz, J.O., et al., *Synopsis of the PI-RADS v2 Guidelines for Multiparametric Prostate Magnetic Resonance Imaging and Recommendations for Use.* Eur Urol, 2016. **69**(1): p. 41-9.

24. Vargas, H.A., et al., *Updated prostate imaging reporting and data system (PIRADS v2) recommendations for the detection of clinically significant prostate cancer using multiparametric MRI: critical evaluation using whole-mount pathology as standard of reference.* Eur Radiol, 2016. **26**(6): p. 1606-12.

25. Weinreb, J.C., et al., *PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2.* Eur Urol, 2016. **69**(1): p. 16-40.

26. Schoots, I.G., et al., *PI-RADS Committee Position on MRI Without Contrast Medium in Biopsy-Naive Men With Suspected Prostate Cancer: Narrative Review.* AJR Am J Roentgenol, 2020: p. 1-17.

27. Radiology, A.C.o. *Prostate Imaging Reporting and Data System v2.1.* 2019; Available from: www.acr.org.

28. Kwak, J.T., et al., *Prostate Cancer: A Correlative Study of Multiparametric MR Imaging and Digital Histopathology.* Radiology, 2017. **285**(1): p. 147-156.

29. Mertan, F.V., et al., *Prospective Evaluation of the Prostate Imaging Reporting and Data System Version 2 for Prostate Cancer Detection.* J Urol, 2016. **196**(3): p. 690-6.

30. Rosenkrantz, A.B., et al., *Proposed Adjustments to PI-RADS Version 2 Decision Rules: Impact on Prostate Cancer Detection.* Radiology, 2017. **283**(1): p. 119-129.

31. Bastian-Jordan, M., *Magnetic resonance imaging of the prostate and targeted biopsy, Comparison of PIRADS and Gleason grading.* J Med Imaging Radiat Oncol, 2018. **62**(2): p. 183-187.

32. Moore, C.M., et al., *Reporting Magnetic Resonance Imaging in Men on Active Surveillance for Prostate Cancer: The PRECISE Recommendations-A Report of a European School of Oncology Task Force.* Eur Urol, 2017. **71**(4): p. 648-655.

33. Gleason, D.F., *Classification of prostatic carcinomas.* 1966(0069-0112 (Print)).

34. Epstein, J.I., et al., *The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma.* Am J Surg Pathol, 2005. **29**(9): p. 1228-42.

35. Epstein, J.I., et al., *The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System.* Am J Surg Pathol, 2016. **40**(2): p. 244-52.

36. Egevad, L., et al., *Prostate cancer grading, time to go back to the future.* BJU Int, 2021. **127**(2): p. 165-168.

37. Sakr, W.A., 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): p. 439-43.

38. Stamey, T.A., *Localized prostate cancer. Relationship of Tumor Volume to clincial significance for treatment of prostate cancer.* Cancer, 1992. **71**: p. 933-938.

39. Ross, H.M., et al., *Do adenocarcinomas of the prostate with Gleason score (GS) </=6 have the potential to metastasize to lymph nodes?* Am J Surg Pathol, 2012. **36**(9): p. 1346-52.

40. Berney, D.M., et al., *Defining clinically significant prostate cancer on the basis of pathological findings.* Histopathology, 2019. **74**(1): p. 135.

41. Epstein, J.I., et al., *Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer.* Jama, 1994. **271**(5): p. 368-74.

42. Edward, S. *NCCN Guidelines version 2.2021 Prostate Cancer.* 2021 [cited 2021; Available from: nccn.org.

43. Wegelin, O., et al., *Comparing Three Different Techniques for Magnetic Resonance Imaging-targeted Prostate Biopsies: A Systematic Review of In-bore versus Magnetic Resonance Imaging-transrectal Ultrasound fusion versus Cognitive Registration. Is There a Preferred Technique?* Eur Urol, 2017. **71**(4): p. 517-531.

44. Yaxley, A.J., et al., *Comparison between target magnetic resonance imaging (MRI) in-gantry and cognitively directed transperineal or transrectal-guided prostate biopsies for Prostate Imaging-Reporting and Data System (PI-RADS) 3-5 MRI lesions.* BJU Int, 2017. **120 Suppl 3**: p. 43-50.

45. Wegelin, O., et al., *The FUTURE Trial: A Multicenter Randomised Controlled Trial on Target Biopsy Techniques Based on Magnetic Resonance Imaging in the Diagnosis of Prostate Cancer in Patients with Prior Negative Biopsies.* Eur Urol, 2019. **75**(4): p. 582-590.

46. Kasivisvanathan, V., et al., *Transperineal magnetic resonance image targeted prostate biopsy versus transperineal template prostate biopsy in the detection of clinically significant prostate cancer.* J Urol, 2013. **189**(3): p. 860-6.

47. Moore, C.M., et al., *Standards of reporting for MRI-targeted biopsy studies (START) of the prostate: recommendations from an International Working Group.* Eur Urol, 2013. **64**(4): p. 544-52.

48. Futterer, J.J., et al., *Can Clinically Significant Prostate Cancer Be Detected with Multiparametric Magnetic Resonance Imaging? A Systematic Review of the Literature.* Eur Urol, 2015. **68**(6): p. 1045-53.

49. Kasivisvanathan, V., et al., *MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis.* N Engl J Med, 2018. **378**(19): p. 1767-1777.

50. Rouvière, O., et al., *Use of prostate systematic and targeted biopsy on the basis of multiparametric MRI in biopsy-naive patients (MRI-FIRST): a prospective, multicentre, paired diagnostic study.* The Lancet Oncology, 2019. **20**(1): p. 100-109.

51. Baco, E., et al., *A Randomized Controlled Trial To Assess and Compare the Outcomes of Two-core Prostate Biopsy Guided by Fused Magnetic Resonance and Transrectal Ultrasound Images and Traditional 12-core Systematic Biopsy.* Eur Urol, 2016. **69**(1): p. 149-56.

52. Porpiglia, F., et al., *Diagnostic Pathway with Multiparametric Magnetic Resonance Imaging Versus Standard Pathway: Results from a Randomized Prospective Study in Biopsy-naive Patients with Suspected Prostate Cancer.* Eur Urol, 2017. **72**(2): p. 282-288.

53. Mottet, N., et al., *EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer-2020 Update. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent.* Eur Urol, 2021. **79**(2): p. 243-262.

54. Drost, F.H., et al., *Prostate MRI, with or without MRI-targeted biopsy, and systematic biopsy for detecting prostate cancer.* Cochrane Database Syst Rev, 2019. **4**: p. CD012663.

55. Ferlay, J., et al., *Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018.* Eur J Cancer, 2018. **103**: p. 356-387.

56. Lundberg, F.E., et al., *Trends in cancer survival in the Nordic countries 1990-2016: the NORDCAN survival studies.* Acta Oncol, 2020. **59**(11): p. 1266-1274.

57. Socialstyrelsen. *Official statistic of Sweden: Cancer incidence in Sweden* 2016 [cited 2016; Available from: www.socialstyrelsen.se.

58. Cancercentrum, *Nationellt vårdprogram prostatacancer.* 2020.

59. Larønningen S, F.J., Bray F, Engholm G, Ervik M, Gulbrandsen J, Hansen HL, Hansen HM, Johannesen TB, Kristensen S, Kristiansen MF, Lam F, Laversanne M, Miettinen J, Mørch LS, Ólafsdóttir E, Óskarsson O, Pejicic S, Petterson D, Skog A, Skovlund CW, Tian H, Toorell N, Virtanen A, Aagnes B, Storm HH, *NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 9.0 (01.03.2021). Association of the Nordic Cancer Registries. Cancer Registry of Norway.* 2021.

60. Engholm, G., et al., *NORDCAN--a Nordic tool for cancer information, planning, quality control and research.* Acta Oncol, 2010. **49**(5): p. 725-36.

61. Draisma, G., et al., *Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context.* J Natl Cancer Inst, 2009. **101**(6): p. 374-83.

62. Draisma, G., 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): p. 868-78.

63. Bangma, C.H., S. Roemeling, and F.H. Schroder, *Overdiagnosis and overtreatment of early detected prostate cancer.* World J Urol, 2007. **25**(1): p. 3-9.

64. Bott, *Anterior prostate cancer: is it more difficult to diagnose?* BJU Int, 2002.

65. Ouzzane, A., et al., *Combined multiparametric MRI and targeted biopsies improve anterior prostate cancer detection, staging, and grading.* Urology, 2011. **78**(6): p. 1356-62.

66. Loeb, S., et al., *Systematic review of complications of prostate biopsy.* Eur Urol, 2013. **64**(6): p. 876-92.

67. Schroder, F.H., et al., *Screening and prostate-cancer mortality in a randomized European study.* N Engl J Med, 2009. **360**(13): p. 1320-8.

68. Hugosson, J., et al., *A 16-yr Follow-up of the European Randomized study of Screening for Prostate Cancer.* Eur Urol, 2019. **76**(1): p. 43-51.

69. Celis-Morales, C.A., et al., *Association between active commuting and incident cardiovascular disease, cancer, and mortality: prospective cohort study.* BMJ, 2017. **357**: p. j1456.

70. Socialstyrelsen. *Screening for prostate cancer*. 2018 [cited 2018; Available from: www.socialstyrelsen.se.

71. RCC, *Rekommendationer om organiserad prostatacancertestning*. 2018.

72. Kohestani, K., et al., *The GOTEBORG prostate cancer screening 2 trial: a prospective, randomised, population-based prostate cancer screening trial with prostate-specific antigen testing followed by magnetic resonance imaging of the prostate*. Scand J Urol, 2021. p 1-9.

73. Cuocolo, R., et al., *Prostate MRI technical parameters standardization: A systematic review on adherence to PI-RADSv2 acquisition protocol*. Eur J Radiol, 2019. **120**: p. 108662.

74. Maier Stephan, W., *Prostate Cancer Diffusion-Weighted Imaging: Does the Choice of Diffusion-Weighting Level Matter?* 2021.

75. Coskun, M., et al., *Impact of bowel preparation with Fleet's enema on prostate MRI quality*. Abdom Radiol (NY), 2020. **45**(12): p. 4252-4259.

76. Wagner, M., et al., *Effect of butylscopolamine on image quality in MRI of the prostate*. Clin Radiol, 2010. **65**(6): p. 460-4.

77. Roethke, M.C., et al., *Prostate magnetic resonance imaging at 3 Tesla: Is administration of hyoscine-N-butyl-bromide mandatory?* World J Radiol, 2013. **5**(7): p. 259-63.

78. Lim, C., et al., *Does a cleansing enema improve image quality of 3T surface coil multiparametric prostate MRI?* J Magn Reson Imaging, 2015. **42**(3): p. 689-97.

79. ACR. *Prostate Imaging Reporting and Data System v2*. 2015; Available from: www.acr.org.

80. Vural, M., et al., *Conspicuity of peripheral zone prostate cancer on computed diffusion-weighted imaging: comparison of cDWI1500, cDWI2000, and cDWI3000*. Biomed Res Int, 2014. **2014**: p. 768291.

81. Maas, *Quantitative Evaluation of Computed High b Value Diffusion-Weighted Magnetic Resonance Imaging of the Prostate*. Invest Radiol, 2013. **48**.

82. Rosenkrantz, A.B., et al., *Computed diffusion-weighted imaging of the prostate at 3 T: impact on image quality and tumour detection*. Eur Radiol, 2013. **23**(11): p. 3170-7.

83. Rosenkrantz, A.B., et al., *The Learning Curve in Prostate MRI Interpretation: Self-Directed Learning Versus Continual Reader Feedback*. AJR Am J Roentgenol, 2017. **208**(3): p. W92-W100.

84. Krishna, S., et al., *Comparison of Prostate Imaging Reporting and Data System versions 1 and 2 for the Detection of Peripheral Zone Gleason Score 3 + 4 = 7 Cancers*. AJR Am J Roentgenol, 2017. **209**(6): p. W365-W373.

85. Becker, A.S., et al., *Direct comparison of PI-RADS version 2 and version 1 regarding interreader agreement and diagnostic accuracy for the detection of clinically significant prostate cancer*. Eur J Radiol, 2017. **94**: p. 58-63.

86. Muller, S., et al., *Poor reproducibility of PIRADS score in two multiparametric MRIs before biopsy in men with elevated PSA*. World J Urol, 2018. **36**(5): p. 687-691.

87. Muller, B.G., et al., *Prostate Cancer: Interobserver Agreement and Accuracy with the Revised Prostate Imaging Reporting and Data System at Multiparametric MR Imaging*. Radiology, 2015. **277**(3): p. 741-50.

88. Rosenkrantz, A.B., et al., *Comparison of interreader reproducibility of the prostate imaging reporting and data system and likert scales for evaluation of multiparametric prostate MRI*. AJR Am J Roentgenol, 2013. **201**(4): p. W612-8.

89. Chernyak, V., et al., *Liver Imaging Reporting and Data System (LI-RADS) Version 2018: Imaging of Hepatocellular Carcinoma in At-Risk Patients*. Radiology, 2018. **289**(3): p. 816-830.

90. Khoo, C.C., et al., *Likert vs PI-RADS v2: a comparison of two radiological scoring systems for detection of clinically significant prostate cancer*. BJU Int, 2020. **125**(1): p. 49-55.

91. Ahmed, H.U., et al., *Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study*. Lancet, 2017. **389**(10071): p. 815-822.

92. Grenabo Bergdahl, A., et al., *Role of Magnetic Resonance Imaging in Prostate Cancer Screening: A Pilot Study Within the Goteborg Randomised Screening Trial*. Eur Urol, 2016. **70**(4): p. 566-573.

93. Ahlberg, M.S., et al., *PCASTt/SPCG-17-a randomised trial of active surveillance in prostate cancer: rationale and design*. BMJ Open, 2019. **9**(8): p. e027860.

94. Borofsky, S., et al., *What Are We Missing? False-Negative Cancers at Multiparametric MR Imaging of the Prostate*. Radiology, 2018. **286**(1): p. 186-195.

95. Palmstedt, E., et al., *Long-term Outcomes for Men in a Prostate Screening Trial with an Initial Benign Prostate Biopsy: A Population-based Cohort*. Eur Urol Oncol, 2019. **2**(6): p. 716-722.

96. Thompson, J.E., et al., *The Diagnostic Performance of Multiparametric Magnetic Resonance Imaging to Detect Significant Prostate Cancer*. J Urol, 2016. **195**(5): p. 1428-1435.

97. Egevad, L., et al., *Standardization of Gleason grading among 337 European pathologists*. Histopathology, 2013. **62**(2): p. 247-56.

98. Proto, A.V., *Radiology 2002--statistical concepts series*. Radiology, 2002. **225**(2): p. 317.

99.   Feinstein, A.R. and D.V. Cicchetti, *High agreement but low kappa: I. The problems of two paradoxes.* J Clin Epidemiol, 1990. **43**(6): p. 543-9.

100.   Chakraborty, D.P., *Recent developments in imaging system assessment methodology, FROC analysis and the search model.* Nucl Instrum Methods Phys Res A, 2011. **648 Supplement 1**: p. S297-S301.

101.   Ahn, S., S.H. Park, and K.H. Lee, *How to demonstrate similarity by using noninferiority and equivalence statistical testing in radiology research.* Radiology, 2013. **267**(2): p. 328-38.

102.   Smith, G.C., et al., *Correcting for optimistic prediction in small data sets.* Am J Epidemiol, 2014. **180**(3): p. 318-24.

103.   Goodman, S.N., *Toward evidence-based medical statistics. 1: The P value fallacy.* Ann Intern Med, 1999. **130**(12): p. 995-1004.

104.   Rosenkrantz, A.B., et al., *Interobserver Reproducibility of the PI-RADS Version 2 Lexicon: A Multicenter Study of Six Experienced Prostate Radiologists.* Radiology, 2016. **280**(3): p. 793-804.

105.   Greer, M.D., et al., *Interreader Variability of Prostate Imaging Reporting and Data System Version 2 in Detecting and Assessing Prostate Cancer Lesions at Prostate MRI.* AJR Am J Roentgenol, 2019: p. 1-8.

106.   Bhayana, R., et al., *PI-RADS versions 2 and 2.1: Interobserver Agreement and Diagnostic Performance in Peripheral and Transition Zone Lesions Among Six Radiologists.* AJR Am J Roentgenol, 2020.

107.   Rosenkrantz, A.B., et al., *Prostate cancer: multiparametric MRI for index lesion localization--a multiple-reader study.* AJR Am J Roentgenol, 2012. **199**(4): p. 830-7.

108.   Niu, X.K., et al., *Diagnostic Performance of Biparametric MRI for Detection of Prostate Cancer: A Systematic Review and Meta-Analysis.* AJR Am J Roentgenol, 2018. **211**(2): p. 369-378.

109.   Kang, Z., et al., *Abbreviated Biparametric Versus Standard Multiparametric MRI for Diagnosis of Prostate Cancer: A Systematic Review and Meta-Analysis.* AJR Am J Roentgenol, 2019. **212**(2): p. 357-365.

110.   Cuocolo, R., et al., *Clinically Significant Prostate Cancer Detection With Biparametric MRI: A Systematic Review and Meta-Analysis.* AJR Am J Roentgenol, 2021. **216**(3): p. 608-621.

111.   Woo, S., et al., *Head-to-Head Comparison Between Biparametric and Multiparametric MRI for the Diagnosis of Prostate Cancer: A Systematic Review and Meta-Analysis.* AJR Am J Roentgenol, 2018. **211**(5): p. W226-W241.

112.   Padhani, A.R., I. Schoots, and G. Villeirs, *Contrast Medium or No Contrast Medium for Prostate Cancer Diagnosis. That Is the Question.* J Magn Reson Imaging, 2021. **53**(1): p. 13-22.

113.   Schoots, I.G., *MRI in early prostate cancer detection: how to manage indeterminate or equivocal PI-RADS 3 lesions?* Transl Androl Urol, 2018. **7**(1): p. 70-82.

114.   Greer, M.D., et al., *Validation of the Dominant Sequence Paradigm and Role of Dynamic Contrast-enhanced Imaging in PI-RADS Version 2.* Radiology, 2017. **285**(3): p. 859-869.

115.   Abreu-Gomez, J., et al., *Dynamic Contrast-Enhanced MRI-Upgraded Prostate ImagingReporting and Data System Version 2 Category 3 Peripheral Zone Observations Stratified by a Size Threshold of 15 mm.* AJR Am J Roentgenol, 2019: p. 1-8.

116.   Druskin, S.C., et al., *Dynamic Contrast Enhanced Magnetic Resonance Imaging Improves Classification of Prostate Lesions: A Study of Pathological Outcomes on Targeted Prostate Biopsy.* J Urol, 2017. **198**(6): p. 1301-1308.

117.   Nordstrom, T., et al., *Prostate-specific antigen (PSA) density in the diagnostic algorithm of prostate cancer.* Prostate Cancer Prostatic Dis, 2018. **21**(1): p. 57-63.

118.   Vellekoop, A., et al., *Population based study of predictors of adverse pathology among candidates for active surveillance with Gleason 6 prostate cancer.* J Urol, 2014. **191**(2): p. 350-7.

119.   Venkitaraman, R., et al., *Predictors of histological disease progression in untreated, localized prostate cancer.* J Urol, 2007. **178**(3 Pt 1): p. 833-7.

120.   Busch, J., et al., *Value of prostate specific antigen density and percent free prostate specific antigen for prostate cancer prognosis.* J Urol, 2012. **188**(6): p. 2165-70.

121.   Gortz, M., et al., *The Value of Prostate-specific Antigen Density for Prostate Imaging-Reporting and Data System 3 Lesions on Multiparametric Magnetic Resonance Imaging: A Strategy to Avoid Unnecessary Prostate Biopsies.* Eur Urol Focus, 2019.

122.   Washino, S., et al., *Combination of prostate imaging reporting and data system (PI-RADS) score and prostate-specific antigen (PSA) density predicts biopsy outcome in prostate biopsy naive patients.* BJU Int, 2017. **119**(2): p. 225-233.

123.   Stavrinides, V., et al., *False Positive Multiparametric Magnetic Resonance Imaging Phenotypes in the Biopsy-naive Prostate: Are They Distinct from Significant Cancer-associated Lesions? Lessons from PROMIS.* Eur Urol, 2021. **79**(1): p. 20-29.

124.   Rais-Bahrami, S., et al., *Natural history of small index lesions suspicious for prostate cancer on multiparametric MRI: recommendations for interval imaging follow-up.* Diagn Interv Radiol, 2014. **20**(4): p. 293-8.

125.   Ghavimi, S., et al., *Natural history of prostatic lesions on serial multiparametric magnetic resonance imaging.* Can Urol Assoc J, 2018. **12**(8): p. 270-275.

126.   Giganti, F., et al., *Natural history of prostate cancer on active surveillance: stratification by MRI using the PRECISE recommendations in a UK cohort.* Eur Radiol, 2021. **31**(3): p. 1644-1655.
127.   O'Connor, L.P., et al., *Changes in Magnetic Resonance Imaging Using the Prostate Cancer Radiologic Estimation of Change in Sequential Evaluation Criteria to Detect Prostate Cancer Progression for Men on Active Surveillance.* European Urology Oncology, 2020.
128.   Bonekamp, D., et al., *Radiomic Machine Learning for Characterization of Prostate Lesions with MRI: Comparison to ADC Values.* Radiology, 2018. **289**(1): p. 128-137.
129.   Chatterjee, A., et al., *Diagnosis of Prostate Cancer with Noninvasive Estimation of Prostate Tissue Composition by Using Hybrid Multidimensional MR Imaging: A Feasibility Study.* Radiology, 2018. **287**(3): p. 864-873.

Paper I