



LUND UNIVERSITY

Real-world data and refinement of MRI in early detection of prostate cancer

Thimansson, Per Erik

2024

Document Version:

Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Thimansson, P. E. (2024). *Real-world data and refinement of MRI in early detection of prostate cancer*. [Doctoral Thesis (compilation), Department of Translational Medicine]. Lund University, Faculty of Medicine.

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Real-world data and refinement of MRI in early detection of prostate cancer

ERIK THIMANSSON M.D.

DEPT. OF TRANSLATIONAL MEDICINE | FACULTY OF MEDICINE | LUND UNIVERSITY





ERIK THIMANSSON serves as consulting radiologist within the Department of Radiology at Helsingborg Hospital.



**FACULTY OF
MEDICINE**

Department of Translational Medicine
Radiology Diagnostics, Malmö

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2024:18
ISBN 978-91-8021-511-4
ISSN 1652-8220



Real-world data and refinement of MRI in early detection of prostate cancer

Real-world data and refinement of MRI in early detection of prostate cancer

Erik Thimansson



LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended at Belfragesalen, BMC, Lund
Friday 9th of February 2024 at 0900

Faculty opponent
Professor Per Liss, Uppsala University, Sweden

Organization: LUND UNIVERSITY
Faculty of Medicine
Department of Translational Medicine
Radiology Diagnostics, Malmö

Document name: Doctoral Dissertation

Date of issue 09-02-24

Author(s): Erik Thimansson

Sponsoring organization:

Title and subtitle: Real-world data and refinement of MRI in early detection of prostate cancer

Abstract:

Background: Prostate MRI (magnetic resonance imaging) plays a critical role in the early diagnostic chain for prostate cancer, placing high demands on the method throughout the process. When this dissertation work began, there was no published research on the implications of introducing commercially available AI-models (artificial intelligence) for prostate MRI into the clinical reality. Likewise, the knowledge was limited regarding cohort-based organized testing for prostate cancer.

Aim: The principal aim of the research presented in this dissertation is to evaluate the refinement of the prostate MRI method through the integration of AI-models and evaluation of DWI (diffusion weighted imaging) in early detection of prostate cancer. Real-world multicenter prostate MRI data are used in the studies. Additionally, it aims to assess the feasibility of an algorithm based on 'MRI first' in a pilot study of organized prostate cancer testing.

Methods: The study cohorts for Papers I-III were based on a common population consisting of 277 consecutive patients who underwent radical robot-assisted prostatectomy at Skåne University Hospital in 2018. For Paper IV a representative cohort of 999 men was created by randomly selecting individuals from 33 municipalities in Region Skåne, using the Swedish population registry.

Paper I presented a retrospective multicenter investigation focusing on a method agreement between today's standard radiologist dependent prostate volume method (ellipsoid formula) compared with an AI-model. The multicenter study presented in Paper II aimed to retrospectively evaluate the performance of prostate whole gland segmentation by two commercially available AI-models. In Paper III we conducted a retrospective cohort study to explore the correlation between apparent diffusion coefficient (ADC) metrics and tumor aggressivity by ISUP grade (International Society of Urological Pathology) in patients undergoing robot-assisted laparoscopic prostatectomy for biopsy-confirmed prostate cancer. Paper IV presents a pilot study that assessed the feasibility of a digitally automated population-based program for organized prostate cancer testing in Region Skåne.

Results: In Paper I The AI-model performed similarly to radiologists in the assessment of prostate volume on MR. The two AI-models evaluated in Paper II perform accurate whole gland prostate segmentation on a par with expert radiologist manual planimetry. In Paper III ADC and ADC ratio did not correlate with tumor aggressiveness defined by ISUP grade in this multicenter MRI study. In Paper IV the proposed model and setup for a community-based, automated, risk-adapted organized prostate cancer testing program was operationally feasible.

Conclusions: The results may contribute to understanding how AI-models can be incorporated into the workflow for efficient patient selection for biopsy based on robust PSA (prostate specific antigen) density values and how AI-models can serve as time-saving tools in a radiologist's preparation for targeted biopsies with MRI/ultrasound fusion technology.

ADC-thresholds should be used with caution in a clinical context and customized validation based on local conditions is recommended.

The pilot study contributes with essential knowledge regarding structuring and implementation of a potential future prostate cancer screening program, offering insights into how risk stratification using PSA density and MRI can help avoid unnecessary biopsies.

Key words: Magnetic resonance imaging, Prostate neoplasms, Deep learning, Prostate Specific Antigen, Diffusion Magnetic Resonance Imaging, Screening
Supplementary bibliographical information

Language English

ISSN and key title: 1652-8220 Lund University, Faculty of Medicine Doctoral Dissertation Series 2024:18

ISBN: 978-91-8021-511-4

Recipient's notes

Number of pages:96

Price

Security classification

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature

Date 2023-12-20

Real-world data and refinement of MRI in early detection of prostate cancer

Erik Thimansson M.D.



LUND
UNIVERSITY

Coverphoto by Midjourney.com/G. Sundström

Copyright pp 1-96 Erik Thimansson M.D.

Paper 1 © by the authors (Open access at Springer Nature)

Paper 2 © by the authors (unpublished manuscript)

Paper 3 © by the authors (Open access at Frontiers in Oncology)

Paper 4 © by the authors (Open access at BJUI International)

Faculty of Medicine

Department of Translational Medicine

Radiology Diagnostics, Malmö

ISBN 978-91-8021-511-4

ISSN 1652-8220

Lund University, Faculty of Medicine Doctoral Dissertation Series 2024:18

Printed in Sweden by Media-Tryck, Lund University

Lund 2024



Media-Tryck is a Nordic Swan Ecolabel
certified provider of printed material.
Read more about our environmental
work at www.mediatryck.lu.se

MADE IN SWEDEN 

Algo significare caris tuis.
Opus facere. Optime facere.
JJ

Contents

List of Papers	10
Papers not included in the thesis	11
Abstract	12
Populärvetenskaplig sammanfattning	14
Thesis at a glance	16
Abbreviations	17
Introduction and aims	19
Prostate cancer	23
The prostate gland	23
Epidemiology	23
Prostate cancer diagnosis	25
Prostate-specific antigen and digital rectal examination.	25
Transrectal ultrasound and biopsies	25
Histologic grading	25
Staging, prognostic risk groups and treatment	27
Prostate MRI	29
Development and sequences	29
MRI protocols	30
Patient preparation	31
Prostate volume and PSA Density	33
Prostate volume	33
Methods for prostate volume assessment	33
Transrectal ultrasound, TRUS	33
MRI	34
Planimetry	35
The complexity of interobserver variation and PSAD	35
Diffusion-weighted prostate MRI and tumor aggressiveness	37
Random movement of protons	37

PI-RADS and DWI.....	39
ADC thresholds and ADC ratios	40
Artificial intelligence in prostate MRI.....	43
Introduction	43
AI-models in Radiology and Prostate MRI.....	45
Image quality	46
Prostate gland segmentation	46
Lesion detection and characterization	48
The challenge with the reference standards, what is really <i>the truth</i> ?.....	49
Organized Prostate Cancer Testing (OPT).....	51
Screening.....	51
Organized testing.....	52
A short introduction to the key statistical methods used in the thesis.....	55
Bland-Altman plots	55
DICE coefficients.....	55
Spearman's rank correlation coefficient	56
Receiver Operating Characteristic (ROC) curves	56
Personal ethical reflections	57
Summary of the studies.....	59
Ethical approvals	59
Cohorts	59
MRI Dataset	61
Methods.....	61
Results	65
Discussion	71
Ad Paper I.....	71
Ad Paper II	71
Ad Paper III	72
Ad Paper IV.....	73
Conclusions	75
Future perspectives	77
Acknowledgements.....	79
References	83

List of Papers

This thesis is based on the following Papers, which will be referred to in the thesis by their Roman numerals, I-IV. The Papers are appended at the end of the thesis.

- I* Deep learning algorithm performs similarly to radiologists in the assessment of prostate volume on MRI.
Thimansson E, Bengtsson J, Baubeta E, Engman J, Flondell-Sité D, Bjartell A, Zackrisson S.
Eur Radiol. 2023 Apr;33(4):2519-2528.
- II* Deep learning performance on MRI prostate gland segmentation: evaluation of two commercially available algorithms compared with an expert radiologist.
Thimansson E, Baubeta E, Engman J, Bjartell A, Zackrisson S.
Manuscript under revision in Journal of Medical Imaging
- III* Correlation between ADC, ADC ratio, and Gleason Grade group in prostate cancer patients undergoing radical prostatectomy: Retrospective multicenter study with different MRI scanners.
Bengtsson J, **Thimansson E**, Baubeta E, Zackrisson S, Sundgren PC, Bjartell A, Flondell-Sité D.
Front Oncol. 2023 Feb 20;13:1079040.
- IV* A pilot study of an organized population-based testing program for prostate cancer.
Alterbeck M, **Thimansson E**, Bengtsson J, Baubeta E, Zackrisson S, Bolejko A, Sandeman K, Carlsson S, Jiborn T, Bjartell A.
BJU Int. 2023 Jul 31. doi: 10.1111/bju.16143. Online ahead of print.PMID: 37523331

Papers not included in the thesis

Designing and Implementing a Population-based Organized Prostate Cancer Testing Program.

Alterbeck M, Järbur E, **Thimansson E**, Wallström J, Bengtsson J, Björk-Eriksson T, Bjartell A, Bratt O, Jiborn T, Arnsrud Godtman R.

Eur Urol Focus. 2022 Nov;8(6):1568-1574. doi: 10.1016/j.euf.2022.06.008. Epub 2022 Jul 8. PMID: 35811285.

Population-based Organized Prostate Cancer Testing: Results from the First Invitation of 50-year-old Men.

Bratt O, Godtman RA, Jiborn T, Wallström J, Akre O, Carlsson S, Nordström T, **Thimansson E**, Alterbeck M, Zackrisson S, Hugosson J, Bjartell A, Lantz A.

Eur Urol. 2023 Dec 1:S0302-2838(23)03272-4. doi: 10.1016/j.eururo.2023.11.013. Epub ahead of print. PMID: 38042646.

Abstract

Background

Prostate MRI (magnetic resonance imaging) plays a critical role in the early diagnostic chain for prostate cancer, placing high demands on the method throughout the process. When this dissertation work began, there was no published research on the implications of introducing commercially available AI-models (artificial intelligence) for prostate MRI into the clinical reality. Likewise, the knowledge was limited regarding cohort-based organized testing for prostate cancer.

Aim

The principal aim of the research presented in this dissertation is to evaluate the refinement of the prostate MRI method through the integration of AI-models and evaluation of DWI (diffusion weighted imaging) in early detection of prostate cancer. Real-world multicenter prostate MRI data are used in the studies. Additionally, it aims to assess the feasibility of an algorithm based on 'MRI first' in a pilot study of organized prostate cancer testing.

Methods

The study cohorts for Papers I-III were based on a common population consisting of 277 consecutive patients who underwent radical robot-assisted prostatectomy at Skåne University Hospital in 2018. For Paper IV a representative cohort of 999 men was created by randomly selecting individuals from 33 municipalities in Region Skåne, using the Swedish population registry.

Paper I presented a retrospective multicenter investigation focusing on a method agreement between today's standard radiologist dependent prostate volume method (ellipsoid formula) compared with an AI-model. The multicenter study presented in Paper II aimed to retrospectively evaluate the performance of prostate whole gland segmentation by two commercially available AI-models. In Paper III we conducted a retrospective cohort study to explore the correlation between apparent diffusion coefficient (ADC) metrics and tumor aggressivity by ISUP grade (International Society of Urological Pathology) in patients undergoing robot-assisted laparoscopic

prostatectomy for biopsy-confirmed prostate cancer. Paper IV presents a pilot study that assessed the feasibility of a digitally automated population-based program for organized prostate cancer testing in Region Skåne.

Results

In Paper I The AI-model performed similarly to radiologists in the assessment of prostate volume on MR. The two AI-models evaluated in Paper II perform accurate whole gland prostate segmentation on a par with expert radiologist manual planimetry. In Paper III ADC and ADC ratio did not correlate with tumor aggressiveness defined by ISUP grade in this multicenter MRI study. In Paper IV the proposed model and setup for a community-based, automated, risk-adapted organized prostate cancer testing program was operationally feasible.

Conclusions

The results may contribute to understanding how AI-models can be incorporated into the workflow for efficient patient selection for biopsy based on robust PSA (prostate specific antigen) density values and how AI-models can serve as time-saving tools in a radiologist's preparation for targeted biopsies with MRI/Ultrasound fusion technology.

ADC-thresholds should be used with caution in a clinical context and customized validation based on local conditions is recommended.

The pilot study contributes with essential knowledge regarding structuring and implementation of a potential future prostate cancer screening program, offering insights into how risk stratification using PSA density and MRI can help avoid unnecessary biopsies.

Populärvetenskaplig sammanfattning

Prostatacancer är den vanligaste cancer hos män i världen och den vanligaste cancer-relaterade dödsorsaken bland män i Sverige. Tidigare har man baserat utredningen av misstänkt prostatacancer på blodprovet PSA (prostata specifikt antigen), men ett förhöjt PSA värde kan ha många olika orsaker förutom cancer. Screening för prostatacancer baserat enbart på PSA resulterar i överdiagnostik (man hittar så kallad icke signifikant cancer, som inte skulle ha skadat mannen under hans livstid). På senare år har magnetkameraundersökning (MR) av prostata fått en allt viktigare roll i utredningen av misstänkt prostatacancer genom paradigmskiftet till ”MR först”, vilket betyder att en MR undersökning skall utföras innan eventuella vävnadsprover tas. De två viktigaste sekvenserna i MR-protokollet är T2 viktad sekvens och Diffusionsviktad sekvens.

MR prostata har två huvudfunktioner när en man utreds för prostatacancer: 1, att hitta eller utesluta förekomst av signifikant cancer och avgöra om tumören är begränsad till körteln eller växer utanför, och 2, att ge en prostatavolym för PSA Densitetsberäkning ($PSA \div \text{prostatavolymen}$). Risken för signifikant cancer i en fokal förändring graderas utifrån en femgradig skala vilken baseras på lokalisation i prostatakörteln i kombination med olika MR-sekvenser där diffusionsviktad sekvens är en viktig komponent. Dagens standardmetoder för volymsbestämning och segmentering (manuell utlinjering av prostatakörteln ytterkonturer) är tidsödande, kräver röntgenläkarens medverkan och uppvisar variation avseende bedömningarna mellan olika röntgenläkare. Segmenteringarna används också för att förbättra träffsäkerheten vid vävnadsprovtagning med hjälp av så kallad fusionsteknik, där MR-bilderna överlagras på dynamiska ultraljudsbilder.

Organiserad testning för prostatacancer (OPT) är regionala projekt rekommenderade av Socialstyrelsen för att få förbättrat underlag inför ett eventuellt framtida införande av befolkningsbaserad screening för prostatacancer. Nyligen har också Europeiska Unionen rekommenderat sina medlemsländer att utvärdera organiserad testning för tidig diagnostik. I denna avhandling belyses ett flertal faktorer som kan få betydelse med hänsyn till den ökade användningen av MR prostata, då behovet av säker och snabb diagnostik är centralt i ett eventuellt framtida screeningprogram.

Studierna är utförda på material som återspeglar den kliniska verkligheten (”real-world data”) och är multi-center studier vilket innebär olika sjukhus och variation

avseende MR-kameror och MR-protokoll, och skiljer sig därmed från merparten av tidigare publikationer inom området.

Arbete I,II och III utgår från en patientgrupp som genomgått MR och radikal prostatektomi (prostata har opererats ut med titthålsteknik och robotassisterad metod) med efterföljande detaljerad analys av vävnadens mikroskopiska struktur. Arbete I är inriktat på utvärdering av AI-modell (artificiell intelligens modell) för prostatavolymbestämning och arbete II är inriktat på utvärdering av två AI-modeller för segmentering av prostatas ytterkontur. I båda dessa arbeten jämförs nya AI-baserade metoder med dagens standardmetoder som kräver röntgenläkarmedverkan. Arbete III undersöker sambandet mellan diffusionsviktad sekvens och allvarlighetsgrad av cancer i ett varierat MR material från många olika MR-kameror. Sambandet mellan ADC-värden (vilka räknas fram från diffusionsviktad sekvens) och tumörens allvarlighetsgrad (Gleason Grade Group) står i fokus. Arbete IV är en pilotstudie inför införandet av organiserad prostatacancertestning i Region Skåne med syfte att utvärdera funktionalitet och utfall i tre olika åldersgrupper.

Sammanfattning av resultat och slutsatser:

Arbete I) AI-modellen visade bättre överensstämmelse och precision än volymbestämningarna utförda av två röntgenläkare användandes dagens standardmetod med manuell mätning.

Arbete II) De två AI-modellerna presterade likvärdigt och i denna multicenterstudie var resultaten i linje med tidigare singlecenter-studier. Tillsammans utgör studie 1 och 2 lovande steg mot att använda AI-modeller för att omfördela röntgenläkarresurser mot mer komplexa arbetsuppgifter än att manuellt mäta prostatavolymer och utföra tidskrävande manuell segmentering.

Arbete III) Till skillnad från tidigare singlecenter-studier kunde inget samband återfinnas mellan ADC och Gleason Grade Group i vår multicenterstudie. Detta manar till försiktighet vid användandet av ADC-trösklar i en klinisk situation och understryker vikten av att harmonisera parametrarna för MRI-sekvenser mellan olika centra.

Arbete IV) OPT-modellen är genomförbar ur ett praktiskt perspektiv. De uppvisade trenderna i denna pilotstudie är viktiga för att optimera planering och bättre kunna förutsäga och dimensionera resursåtgången för ett eventuellt framtida screeningprogram.

Thesis at a glance

Study	Question	Method	Results and Conclusions
I	Can a commercially available AI-model assess prostate volume on MRI on par with today's gold standard radiologist dependent method on a clinical dataset?	Retrospective multicenter method agreement study comparing radiologist ellipsoid formula with AI-model, expert radiologist manual planimetry and specimen weight volume as reference standards.	The AI-model performs similarly to radiologists in the assessment of prostate volume on MR. This could potentially liberate the radiologist for more complex tasks.
II	What is the performance level of two commercially available AI-models for prostate gland segmentation on a real-life diverse dataset?	Retrospective multi-scanner study comparing the performance of two AI-models with expert radiologist manual planimetry as the reference standard	The two AI-models perform accurate whole gland prostate segmentation on a par with expert radiologist manual planimetry. AI-models have the potential to facilitate the radiologist's preparatory work for fusion biopsies.
III	Is there a correlation in a heterogeneous MRI dataset between ADC, ADC ratio, and Gleason Grade group in prostate cancer patients undergoing radical prostatectomy?	Retrospective consecutive cohort study exploring the correlation between ADC metrics and ISUP grade in preoperative prostate MRI in patients undergoing RALP for biopsy-confirmed prostate cancer.	ADC and ADC ratio did not correlate with tumor aggressiveness defined by ISUP grade in this multicenter MRI study. The result of this study is opposite to previous research in the field.
IV	What is the feasibility of a digitally automated population-based program for organized prostate cancer testing (OPT) in Southern Sweden?	999 randomly selected men aged 50, 56, or 62 years were invited. Multiple steps in the OPT process were performed by an automated digital management system. Risk stratification was based on PSA level, PSAD, and prostate MRI.	The proposed model and setup for a community-based, automated, risk-adapted OPT program was operationally feasible. Increasing the age for invited men may have impact on participation rates, PSA levels, and prostate cancer detection rates.

Abbreviations

ADC	Apparent Diffusion Coefficient
AI	Artificial Intelligence
DRE	Digital Rectal Examination
DL	Deep Learning
MRI	Magnetic Resonance Imaging
DWI	Diffusion-Weighted Imaging
DCE	Dynamic Contrast Enhanced
PI-RADS	Prostate Imaging Reporting and Data System
PC	Prostate Cancer
PSA	Prostate-Specific Antigen
PSAD	Prostate-Specific Antigen Density
TRUS	Transrectal Ultrasound
T2W	T2 weighted
T1W	T1 weighted
OPT	Organized Prostate cancer Testing

Introduction and aims

Prostate cancer (PC) is the most common cancer in men worldwide[1]. In Sweden, prostate cancer is the most common cancer among men with around 10 000 new cases annually[2]. It is also the most frequent cause of cancer-related deaths among men in Sweden where approximately 5% of Swedish men succumb to prostate cancer[2]. Prostate cancer is a heterogeneous disease ranging from indolent non-significant cancer to highly aggressive early metastasizing cancer.

The timing of both diagnosis and treatment is of the utmost importance. We aim to detect prostate cancer before it has advanced too far to be curable, while also avoiding overdiagnosis and overtreatment of indolent cancer.

Magnetic Resonance Imaging (MRI) of the prostate gives excellent morphological information from T2-weighted images (T2W) and biological information from diffusion-weighted images (DWI) and is the preferred radiological examination to detect cancer. The paradigm shift to 'MRI first', meaning that an MRI examination is performed before biopsies, changed the diagnostic work-up for PC; a positive MRI with cancer suspicion stratifies men to undergo biopsies, and a negative MRI without cancer suspicion will stratify men to be spared from biopsy[3-5]. This pathway potentially results in less overdiagnosis of indolent cancer, less underdiagnosis of significant cancer, fewer biopsies performed and thereby fewer biopsy-related serious infections like sepsis.

Screening for PC is attractive due to a long, organ-confined, asymptomatic stage in contrast to an often incurable disease when symptomatic. Prostate Specific Antigen (PSA) based screening trials have shown reduced cancer-specific mortality but have challenges with overdiagnosis of indolent cancers[6]. Screening with PSA in combination with MRI may decrease the problem with overdiagnosis of indolent cancers[7]. Recently, the European Union Council recommended a stepwise approach to PC screening, including pilot studies and further research[8]. The Swedish National Board of Health and Welfare in 2018 recommended against a national screening program for prostate cancer but encouraged organized PSA testing[9].

The demand for prostate MRI is expected to increase in the future due to continuous adaptation to the 'MRI first' pathway and the implementation of organized testing. Furthermore, the introduction of national screening programs for prostate cancer would significantly boost the demand for prostate MRI. Currently, there is already

a shortage in manpower (radiologists specialized in prostate MRI) and a potential solution to address the challenge of the increased workload is the utilization of the software-assisted MRI review. Artificial Intelligence (AI) models have been developed to aid in MRI analysis, covering tasks such as prostate volume calculations and the detection and characterization of suspicious cancer lesions. Among important requirements for such AI-models are predictability, robustness, and generalizability.

Prostate MRI plays a critical role in the early diagnostic chain for prostate cancer, placing high demands on the method throughout the process. Examination protocols, review, interpretation, and reporting must be structured and standardized, and feedback from biopsy results to the radiologist must be facilitated.

When this dissertation work began, there was no published research on the implications of introducing commercially available AI-models into our clinical reality, where examinations are conducted in various departments using MRI scanners of different field strengths and models. Our approach to addressing this knowledge gap was to gradually start testing and validating AI-models for tasks that are currently important but also time-consuming and tedious for radiologists (in Paper I, manual measurement of prostate volume, and in Paper II, manual outlining of prostate contours). AI-models that detect and characterize cancer-suspected lesions rely heavily on information from the diffusion-weighted sequence, and therefore, we aimed to better understand how to use ADC values from the DWI series. In Paper III, we evaluated how ADC values correlate with the severity of cancer in our clinical reality.

Likewise, when the dissertation work commenced, the knowledge landscape was limited regarding cohort-based organized testing for prostate cancer. In the pilot study in Paper IV, we assessed the feasibility and outcomes of organized testing in three different age groups invited to prostate cancer testing.

The principal aim of the research presented in this thesis is to evaluate the refinement of the prostate MRI method through the integration of AI-models and evaluation of DWI in early detection of prostate cancer. Additionally, it aims to assess the feasibility of an algorithm based on 'MRI first' in a pilot study of organized prostate cancer testing.

The specific aims of the studies presented in the Papers were:

- I.) to compare a commercially available AI-model and today's radiologist dependent standard method for prostate volume assessment.
- II.) to compare prostate contour assessment by two commercially available deep learning algorithms with those of an expert radiologist.

III.) to explore the correlation between absolute ADC values and ADC ratios with ISUP grade in a diverse patient cohort undergoing robot-assisted laparoscopic prostatectomy (RALP) imaged with various MRI scanners.

IV.) to describe participation rates, PSA and MRI outcomes, and prostate cancer incidence in men aged 50, 56, and 62 years in the pilot project of organized prostate cancer testing in southern Sweden.

Prostate cancer

The prostate gland

Early descriptions of prostate anatomy date back as far as the mid-16th century when Andreas Vesalius published his observations of the male accessory glands[10]. In the early 1900s, studies on embryology were published[11, 12]. The gland is situated low in the pelvic cavity, located beneath the urinary bladder, encircling the prostatic urethra, and positioned just anterior to the rectum, collectively constituting a surgically challenging position. The main function of the prostate, is to generate a conducive and favourable chemical environment for the transit of sperm. The prostate gland is anatomically divided into different zones, the peripheral zone and the transition zone being the two most important. The urethra traverses the prostate, facilitating the passage of semen and urine from the ejaculatory ducts and bladder, respectively. Seminal vesicles, located on each side of the prostate, contribute the majority of the fluid in semen. Prostate cancer most commonly originates in the peripheral zone, accounting for approximately 70% of cases. The remaining cancers are distributed in the transition zone (10-20%) and central zone (5-10%)[13].

Epidemiology

Prostate cancer is the most common cancer in men worldwide[1]. Its incidence and mortality rates are highest in developed countries like North America and Europe. In Sweden, prostate cancer is the most common cancer among men with around 10 000 new cases annually[2]. The number of new cases of prostate cancer increased rapidly around the turn of the century, doubling from 1990 to 2004 (Figure 1). The primary reasons were increased diagnostic activity through PSA testing and systematic prostate biopsies, but the growing number of elderly men in the population also contributed. In 2021, approximately 125,000 men were living with diagnosed prostate cancer in Sweden[14, 15], which is three times more than 20 years ago. The increase is due in part to the growing number of elderly men in the population, as well as the fact that men are being diagnosed at an earlier stage today compared to the past, and men with advanced prostate cancer are living longer due to better treatment. It is also the most frequent cause of cancer-related deaths among

men in Sweden where approximately 5% of Swedish men succumb to prostate cancer[2]. The age-standardized mortality for prostate cancer decreased slightly between 1970 and 2020, with a declining trend since the early 2000s (Figure 1).

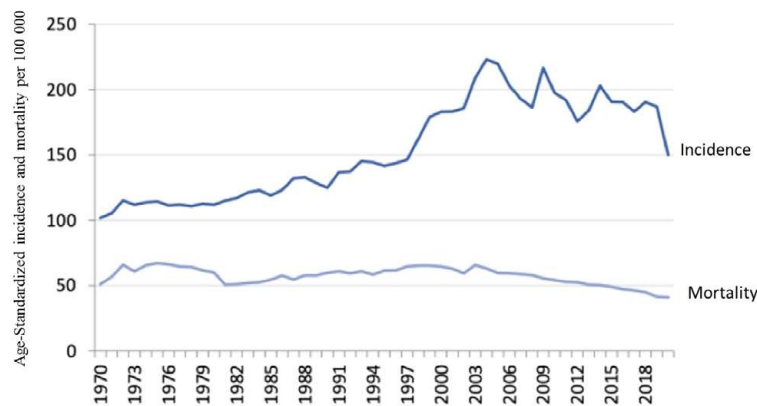


Figure 1. Age standardized prostate cancer incidence and mortality in Sweden 1970- 2020. Data from NORDCAN[14].

The uneven occurrence of the disease suggests that our Western lifestyle increases the likelihood of prostate cancer. Obesity[16] and a high consumption of dairy products[17] have been recognized as likely risk factors. Chronic inflammation may play a role as an etiological factor, although the exact causal connection remains unclear. Some evidence suggests that tobacco smoking increases the risk of men developing aggressive prostate cancer[18].

Heredity plays a substantial role in the risk of developing prostate cancer. Sons and brothers of individuals with prostate cancer face 1.5–3 times higher risks, particularly if multiple family members are affected. Early-onset cases increase mortality risk[19]. Maternal lineage alone does not significantly increase risk unless coupled with breast or ovarian cancer history. If a brother is affected, the risk increase is similar to paternal heredity. The genetic basis for hereditary susceptibility to prostate cancer is intricate. Two genes with clear clinical relevance have been identified, BRCA2 and HOXB13, and mutations in these genes confer an increased risk of prostate cancer, particularly with respect to BRCA2[20, 21], where the cancer presents at an early age and exhibits a more aggressive nature.

The natural progression of prostate cancer varies among individuals. Observational studies have shown that many localized tumors grow slowly and have a minimal risk of symptoms or mortality[22-24]. These indolent prostate cancers are prevalent, especially in men over the age of seventy[25, 26]. In contrast, other tumors are aggressive and can significantly impact long-term survival.

Prostate cancer diagnosis

For patients with clinical suspicion of prostate cancer and elevated PSA the standard pathway, according to 'MRI first' is to refer the patient for a prostate MRI.

Prostate-specific antigen and digital rectal examination.

Prostate Specific Antigen (PSA) is an antigen generated by both normal and malignant cells in the prostate gland. Although PSA is prostate-specific, it is not prostate cancer-specific; PSA levels also rise in benign prostatic hyperplasia and inflammatory processes. Hence, various ratios have been explored, such as free/total PSA and the prostate health index[27, 28]. In current guidelines, the ratio of PSA to prostate volume, PSA density (PSAD)[29-31], is the primary approach, often guiding clinical decision-making, as seen in the European[32] and Swedish National Clinical Cancer Care guidelines for prostate cancer[33]. Studies indicate that approximately 20% of men with normal PSA levels actually have PC, while many men with elevated PSA levels do not[34, 35]. Given that the majority of prostate cancer lesions are situated in the peripheral zone, digital rectal examination (DRE) proves effective in detecting lesions of a certain size within this region. However, DRE is unable to reach lesions in other zones[36].

Transrectal ultrasound and biopsies

Transrectal ultrasound (TRUS) and needle-core biopsy play crucial roles in the investigation of prostatic disease. TRUS is routinely utilized to measure prostate size, evaluate prostatic anatomy to some extent, and detect suspicious tumor areas. If the initial prostate cancer (PC) investigation, involving PSA, DRE, and MRI, raises suspicion of PC, needle-core biopsies are conducted using TRUS after local anesthesia. In the traditional diagnostic pathway, when PC is suspected based on PSA and DRE, systematic biopsies are taken from 10-12 standardized locations in the prostate. However, current guidelines from the European Association of Urology[32] and the Swedish national guidelines for PC now recommend a pre-biopsy MRI[33]. Based on the MRI results, along with those of PSA and DRE, targeted and/or systemic biopsies are then performed, reflecting a shift in the diagnostic approach.

Prostate MRI is described in more detail in separate chapters.

Histologic grading

The Gleason grading system (Figure 2), developed by Donald F. Gleason in 1966[37], is used worldwide by pathologists to grade prostatic adenocarcinomas

based on the degree of tissue abnormality[38]. It involves identifying five different histologic patterns and scoring the growth pattern from well to poorly differentiated.

The pathology report encompasses two patterns that are combined to determine the Gleason score, such as $3+3 = 6$ for low-risk cancer. Initially, the sum was calculated by considering the two most prevalent patterns. However, in 2005, the International Society of Urological Pathology (ISUP) consensus meeting revised the definition to consistently include the highest grade pattern when evaluating biopsy cores. This alteration entailed incorporating both the most common pattern and the highest grade pattern found in any of the cores[39]. In the year 2014, a novel ISUP consensus meeting took place, wherein the idea of grade groups for the communication of biopsy findings, commonly referred to as ISUP grade groups, was introduced[40]. These ISUP grade groups are designated with numbers ranging from 1 to 5. ISUP 1 corresponds to a Gleason score of $3+3$, ISUP 2 to a Gleason score of $3+4$, ISUP 3 to a Gleason score of $4+3$. ISUP 4 corresponds to any Gleason sum of 8, and ISUP 5 represents any Gleason sum ranging from 9 to 10.

However, a limitation of histological grading is the variability among pathologists, with differences observed between general pathologists and experts in uropathology[41, 42] but also between experts[43]. This variability is significant and can impact treatment decisions, as histological grade is a strong prognostic factor.

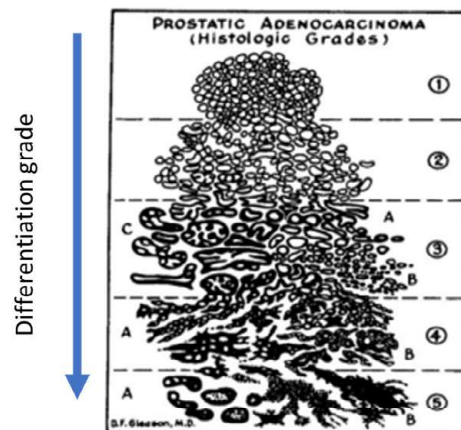


Figure 2. Gleason's five patterns: 1 = Small, uniform glands; 2 = Increased stroma between glands; 3 = Infiltrative margins; 4 = Neoplastic glands forming irregular masses; 5 = Glands only formed occasionally. Adapted from Gleason[37].

Staging, prognostic risk groups and treatment

The staging of PC follows the Tumor, Node, Metastasis (TNM) classification[38]. The clinical tumor stage (cT-stage) is determined by DRE findings, providing insight into the primary cancer's extent. Notably, imaging techniques such as MRI are not taken into account in the TNM classification. The pathologic tumor stage (pT-stage) relies on the pathology report post-radical prostatectomy and aligns closely with the cT-stage, with the distinction that all PC identified during pathology examination are categorized as at least stage T2, and no further sub-stages are acknowledged for T2 tumors. Node-stage (N) and metastasis-stage (M) indicate whether and to what extent the disease has spread to lymph nodes (N) and other parts of the body (M). N- and M-stages are determined through imaging methods like bone scans and increasingly with prostate-specific antigen (PSMA) Positron Emission Tomography techniques.

The choice of treatment for prostate cancer depends on the risk group, the patient's overall health and life expectancy, and patient preference. Untreated prostate cancer that does not show signs of metastasis can be classified into various prognostic risk groups. These groups are determined based on the PSA level, clinical tumor stage (cT-stage), and Gleason score obtained from systematic biopsy (there is a lack of evidence-based risk group categorization for prostate cancer diagnosed with targeted biopsies). These definitions were established by the D'Amico risk classification system in 1998[44], which categorizes tumors into low, intermediate, and high-risk groups. Over time, this risk classification system has undergone modifications. The current Swedish national guidelines for PC further divide the low-risk group into very low and low-risk subgroups[33]. Localized prostate cancer is potentially curable, while advanced cases are considered incurable and treated palliatively. Metastasis investigations are typically performed for high-risk tumors, while low and intermediate-risk groups may not require further investigation. Treatment options for incurable cases include expectance, hormonal treatment, chemotherapy, radiotherapy, and transurethral resection. The most suitable treatment strategy is determined based on the disease stage, progression rate, and symptoms. The patient's remaining life expectancy is important in determining treatment, with watchful waiting being recommended for older patients with a short life expectancy[32, 45] and active surveillance[32, 33], radiotherapy[46, 47], or surgery[24] being options for patients with a longer life expectancy.

Prostate MRI

Development and sequences

Since the pioneering prostate MRI studies of the 1980s[48], this imaging modality has witnessed extraordinary advancements. Initially, the primary focus was on harnessing the anatomical and morphological resolutions offered by T2-weighted sequences (t2W) to stage prostate tumors. The introduction of endorectal coils facilitated enhanced resolution and signal-to-noise ratios[49, 50], albeit with challenges related to the coil being uncomfortable for the patient and to prostate deformation during assessment. Endorectal coils are not used in Sweden today. Another notable development was the incorporation of MRI spectroscopy, allowing for the measurement and analysis of the chemical composition of prostate tissue[51]. However, this examination proved time-consuming both in terms of acquisition and review and is now infrequently conducted.

In recent years, diffusion-weighted imaging (DWI) sequences have emerged as a cornerstone of the prostate MRI protocol. This imaging approach capitalizes on the principle that water diffusion within tissues is restricted or hindered by cellular structures, such as cell membranes and tissue organization. DWI in prostate MRI has shown remarkable sensitivity in detecting prostate cancer, owing to the limited water diffusion typically observed in malignant tissues due to their unique cellular structure[52, 53]. The evolution of technology has led to increased gradient strength in modern MR scanners, resulting in significantly more sophisticated diffusion weighted images. DWI is described in more detail in the following chapters.

One can liken the diffusion series to illuminating small light bulbs that aid radiologists in identifying pathological features within the images. It is important to note that these 'light bulbs' are not exclusive to cancer; several non-neoplastic conditions also manifest restricted diffusion, with inflammatory processes being the most common source of interpretational challenges[54].

In addition to T2W and DWI sequences, dynamic contrast-enhanced (DCE) imaging has become a standard recommendation in international guidelines[55]. DCE involves sequential scans of the prostate at short intervals (temporal resolution > 15 seconds for a duration of ≥ 2 minutes) to provide insights into how tumors enhance with and wash out contrast[56].

T1-weighted (T1W) sequence is still recommended[55], but its relevance has diminished over time[57]. The main objective of utilizing T1W images is to ascertain the existence of hemorrhage in the prostate and seminal vesicles[55, 58]. Nowadays, it is less common for patients to have undergone biopsies before their initial MR examination, resulting in a reduced utility for T1W sequences. Nonetheless, our institution retains a simplified T1W imaging sequence in our protocol (scan time <30 seconds), primarily due to its broader field of view, offering additional insights into enlarged lymph nodes and bone marrow abnormalities.

MRI protocols

In 2012, the European Society of Urogenital Radiology (ESUR) developed guidelines, the Prostate Imaging-Reporting and Data System (PI-RADS), to establish a standardized approach for performing and interpreting prostate MRI. These guidelines were subsequently revised to version 2 in 2015 through a collaborative effort between ESUR and the American College of Radiology. In 2019, a minor update (PI-RADSV2.1) was made to the guidelines[55].

A multiparametric prostate MRI protocol typically incorporates T2W, DWI, and DCE sequences. At the national level, there is a consensus to recommend a biparametric protocol (including T2 and DWI)(Figure 3) for the evaluation of suspected prostate cancer[33]. This recommendation is based on the consideration that the drawbacks associated with contrast administration, such as intravenous catheter placement, the risk of allergic reactions, unclear long-term effects of gadolinium deposition in tissues, increased costs, and extended examination times, outweigh the incremental value provided by DCE.

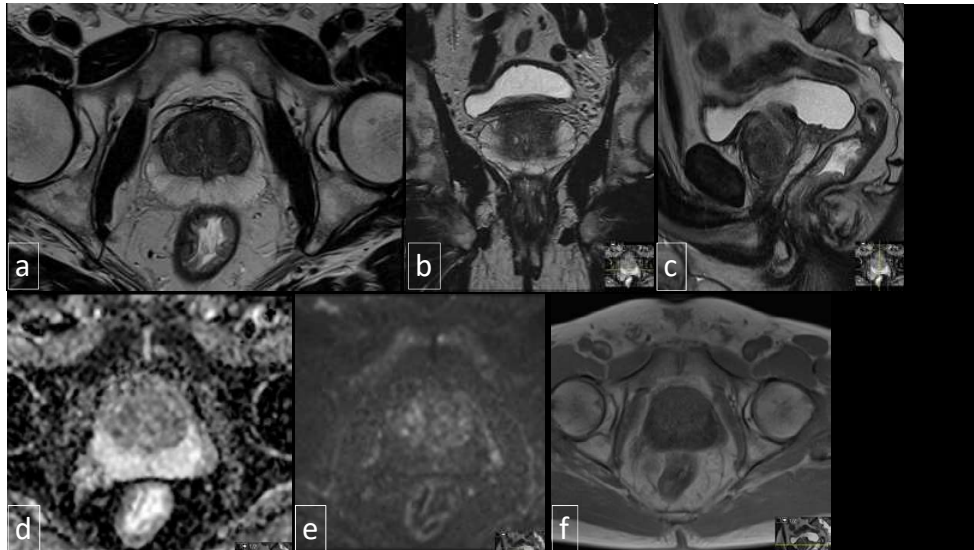


Figure 3. Normal prostate MRI at 1.5 T scanner at our institution. Upper row: T2W transversal (a), coronal (b) and sagittal (c) images. Lower row: Apparent Diffusion Coefficient (ADC) map (d), b 1500 calculated (e) and T1W (f). Collage Dr Thimansson.

Patient preparation

The presence of an empty rectum is a key factor in determining the quality of prostate MRI images[55, 59]. The degree of rectal distention has a negative correlation with image quality on T2W and DWI. This is due to the motion artifact caused by rectal peristalsis and the susceptibility artifact caused by gas[60]. PI-RADS recommends using minimal rectal preparation[55], while other authorities also advocate rectal preparation through the use of a cleansing enema, which could provide benefits[60-62]. It is suggested that the use of an antispasmodic agent may be beneficial for certain patients[55]. However, studies examining the effects of rectal preparation on prostate MRI have yielded conflicting results[63-65]. A comprehensive evidence-based review conducted in 2021 revealed significant variability in the available data and emphasized the need for higher quality studies[66, 67].

Prostate volume and PSA Density

Prostate volume

Prostate volume typically increases with age due to benign prostatic hyperplasia, where the transition zone enlarges due to the development of hypertrophic nodules[68, 69]. It is also common to observe the growth of a third lobe, known as the lobus tertius or median lobe hypertrophy, in the base of the prostate. This protrudes into the bladder and can act as a ball valve, obstructing the upper sphincter and causing lower urinary tract symptoms[70]. Treatment decisions for prostate cancer may be influenced by the size of the prostate, as there is reluctance to irradiate significantly enlarged prostates, both regarding brachytherapy and external beam radiation treatments[71-74]. However, volume is most crucial when correlated with the patient's current PSA value[75, 76].

Methods for prostate volume assessment

Previously, prostate size was estimated through DRE, which provided uncertain estimations[77]. The prostate resembles a truncated cone in shape, constituting an imperfect geometric body. However, simplifying the organ as an ellipsoid or a bullet, both formulas for the volume of these bodies have been employed[78]. The ellipsoid formula, expressed as width x depth x height x ($\pi/6$), has become dominant and is recommended by PI-RADS[55]. Idealizing the prostate as a perfect ellipsoid simplifies but introduces sources of error[79-81], such as in median lobe hypertrophy[82], where there is a risk of volume overestimation (see Figure 4).

Transrectal ultrasound, TRUS

Ultrasound has been, and still is, the predominant method for determining prostate volume. While a transabdominal approach has been described[83], the prevailing method involves a transrectal approach[84], allowing proximity to the target organ for measurement. The method's advantages include being cost-effective, widely available among urologists, rapid, and relatively robust, while disadvantages include sensitivity to artifacts, difficulty measuring significantly enlarged prostates,

and patient discomfort[85, 86]. Studies have also demonstrated the feasibility of estimating prostate volume with computed tomography (CT), and software-assisted volume calculations for CT have been explored[87].

MRI

MRI for volume calculation has become increasingly crucial, especially as more men undergo prostate MRI due to the paradigm shift toward 'MRI first', and importantly, with the rise in organized prostate cancer testing [88]. In our department's OPT, a man with elevated PSA $>3 \mu\text{g/l}$ is offered an prostate MRI before a urological consultation implying that the volume obtained through MRI is the only known volume. Subsequent management is entirely dependent on the MRI outcome. Prostate volume calculation on MRI is fundamental as it forms the basis for the patient's PSAD. Whether a man should be recommended for biopsy is determined by PSAD and the PI-RADS score obtained, with further management dictated by either/both parameters or their combination[33, 89].

The prevailing method, also endorsed in national[33] (RCC) and international[55] guidelines, is the ellipsoid formula. In this method, the radiologist manually places six measurement points on T2W images. Prostate width is measured on transverse images, and depth and height are measured on sagittal images (Figure 4). The advantages of the method include its demonstrated robustness, reasonably predictable outcomes, and avoidance of some challenges faced by TRUS (such as artifacts and difficulties in delineating significantly enlarged prostates). Method drawbacks include being time-consuming and tedious for the radiologist, exhibiting inter-observer variation, the formula assuming the prostate is a perfect ellipsoid, and being less suitable for cases post-transurethral resection of the prostate.

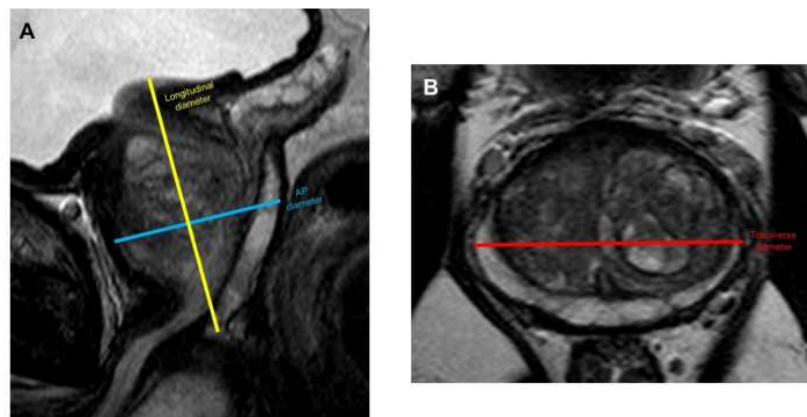


Figure 4. Prostate volume assessment as recommended by PI-RADS using the ellipsoid formula. From the PI-RADS document v 2.1[55].

Planimetry

Hence, methods for prostate volume calculation via MRI have been explored to determine their potential for greater robustness, reduced variation, avoidance of sources of error, more predictable output, and a volume closer to the true prostate volume from reference standard surgical specimens obtained from radical prostatectomy[78, 90-92]. The method identified as most suitable in this context is segmentation or planimetry[93, 94]. This involves outlining the prostate's surface contour on high-resolution T2-weighted images, typically on transverse sections, followed by merging these outlines to obtain a software-assisted volume calculation. The significant drawback is its time-consuming and tedious nature for the radiologist, requiring execution in external software beyond a routine workflow, adding complexity, increased cost, and potential sources of error. On the positive side, segmentation serves purposes beyond volumetric assessment. It is integral to pre-work for fusion biopsies where, in addition to the outer contour, suspected cancer lesions and often the urethra are outlined. A process called re-slicing orients these outlines to mimic the probe's position during TRUS[95]. Geometric coordinates created are exported to the ultrasound machine used by urologists for biopsy orientation. This step involves overlaying the prostate's outer contour and cancer-suspected lesions from MRI onto live ultrasound images, a process termed 'fusion', thus the name 'fusion biopsy'. Fusion biopsies are gaining popularity and are believed to enhance accuracy, particularly in challenging areas of the prostate, especially for small lesions; such as those located ventrally, deep apically, far laterally, and high up in the base[31]. Segmentation is also employed in external beam radiation therapy and brachytherapy planning.

There has been considerable interest in automating segmentations with software-assisted methods, and in the last decade, the development of AI-models based on deep learning and convolutional neural networks has seen a breakthrough.

The complexity of interobserver variation and PSAD

A challenge lies in inter-observer variation with manual measurement methods, and no method is immune to these downsides[81, 96]. In a recent national quality study on prostate MRI assessment conducted by the national consortium for research in prostate MRI (in manuscript), in which I have taken part, interesting results emerged, revealing significant variations in prostate assessment using the ellipsoid formula. Patterns of outliers were also observed, where the same individual exhibited systematic underestimations of prostate volume. Therefore, uncertainties pile up when using methods with measurement inaccuracies and inter-observer variations. To further complicate matters, there is a non-negligible uncertainty in the denominator of the PSAD ratio. In a recently conducted study[97] by the

Swedish consortium for research in organized prostate cancer (SCROPT), we found significant variations in PSA analysis equipment among the three compared health regions. This is crucial as many men hover near the PSA threshold of 3 $\mu\text{g/l}$ for MRI referral, leading to a significantly higher number of MRI's in the region with the most generous PSA measurement equipment in our study. It is essential to note that historically, PSAD thresholds are calculated based on data from TRUS volumes. As we increasingly obtain volumes from MRI, those volumes are probably more accurate[91], but PSAD thresholds may need adjustment based on this fact.

Overall, this calls for humility regarding the rigidity of PSAD threshold values[98]. We will need to adjust PSAD thresholds and their significance in the decision tree as we gather more data and knowledge about the various tools in the toolbox.

Diffusion-weighted prostate MRI and tumor aggressiveness

Random movement of protons

In malignant tumors, there is typically rapid growth, leading to a deviation from the normal tissue architecture at the affected site. Tumors exhibit higher cell density and reduced associated loose tissue in the local environment. In the prostate, a common scenario is the occurrence of a tumor in the peripheral zone, which accounts for about 3/4 of all tumors. In a healthy individual, the peripheral zone comprises atrophied glandular structures characterized by low cell density and preserved interstitium, creating intercellular spaces where water molecules can freely diffuse in random patterns.

When a typical acinar adenocarcinoma develops in the peripheral zone, it is composed of densely packed glandular structures that deviate from the normal glands to varying degrees depending on the tumor's level of differentiation. The tumor cells are tightly packed, displacing the intercellular compartment, which restricts the free diffusion of water molecules—a phenomenon referred to as restricted diffusion[99]. Figure 5 illustrates two examples (a, tumor map from a surgically removed prostate with a unilateral tumor and b, targeted biopsy core). Even without high magnification, the noticeable difference in cell density between the tumor and the preserved prostate tissue is apparent.



Fig 5. a, tumor map from a surgically removed prostate with a unilateral tumor and b, targeted biopsy core. Even without high magnification, the noticeable difference in cell density between the tumor and the preserved prostate tissue is apparent. Digitized pathology images from Region Skåne.

The Apparent Diffusion Coefficient (ADC) is a quantitative metric derived from various b-values within a diffusion series[100, 101]. The b-value signifies the strength of magnetic gradients applied during imaging, influencing the speed and direction of water molecule movement. It is measured in seconds per square meter (s/m²). Lower b-values (typically between 0 and 100 s/m²) indicate predominantly non-directional diffusion of water molecules, while higher b-values (usually

between 500 and 1000 s/m²) result in more extensive movement of water molecules in specific directions.

PI-RADS and DWI

In the PI-RADS system[55], DWI holds paramount importance in the peripheral zone (Table 1) and is the second most crucial in the transition zone (where T2W is the dominant sequence). Radiologists assign PI-RADS scores from 1 to 5 to focal lesions, with higher scores indicating an increased risk of significant cancer. A PI-RADS score of 1 signifies a very low risk, suggesting the unlikelihood of clinically significant cancer. PI-RADS 3 represents an intermediate risk, where the presence of clinically significant cancer is uncertain. PI-RADS 5 corresponds to a very high risk, indicating a high likelihood of clinically significant cancer.

Table 1 a outlines how PI-RADS scores are determined in the peripheral zone.

DWI	T2W	DCE	PI-RADS
1	Any*	Any	1
2	Any	Any	2
3	Any	-	3
		+	4
4	Any	Any	4
5	Any	Any	5

* "Any" indicates 1-5.

Table 1 b. shows how PI-RADS scores are determined in the peripheral zone.

Score	Peripheral Zone (PZ) or Transition Zone (TZ)
1	No abnormality (i.e., normal) on ADC and high b-value DWI.
2	Linear/wedge shaped hypointense on ADC and/or linear/wedge shaped hyperintense on high b-value DWI.
3	Focal (discrete and different from background) hypointense on ADC and/or focal hyperintense on high b-value DWI; may be markedly hypointense on ADC or markedly hyperintense on high b-value DWI, but not both.
4	Focal markedly hypointense on ADC and markedly hyperintense on high b-value DWI; <1.5 cm in greatest dimension.
5	Same as 4 but ≥1.5 cm in greatest dimension or definite extraprostatic extension/invasive behavior.

ADC thresholds and ADC ratios

As shown in the aforementioned discussion on tumor cellularity, there is an inverse correlation between ADC values and cellularity, as well as between ADC values and malignant tumors[102-106]. This relationship also applies to prostate cancer and forms the background for Paper II in this dissertation. The association is stronger in the peripheral zone and weaker in the organized chaos with naturally occurring denser cellular structures found in the transition zone[107, 108].

PI-RADS does not stipulate absolute threshold values but provides advisory formulations as shown in the table 1b above. This allows for some subjectivity[109], and we have observed that this subjectivity can lead to overdiagnosis if different scanners at the same institution have ADC maps with varying appearances, where focal hypointense lesions can be benign on one scanner and malignant on another.

Multiple independent studies have demonstrated an inverse correlation between ADC values and ISUP grade[110-115], but no consensus has been reached regarding definitive threshold values[112, 116, 117]. Regarding ADC values for differentiating between benign and malignant lesions, the PI-RADS document stipulates: "qualitative visual assessment is often used as the primary method to assess ADC. Nevertheless, ADC values, using a threshold of 750-900 $\mu\text{m}^2/\text{sec}$, may assist differentiation between benign and malignant prostate tissues, with ADC values below the threshold correlating with clinically significant cancers". In a literature review by Meyer comprising 1633 cases published in 2020[118], the pooled mean for non-significant lesions (Gleason 5 and 6) was 1100 $\mu\text{m}^2/\text{sec}$, while for significant cancer (Gleason ≥ 7) it was 860 $\mu\text{m}^2/\text{sec}$. A threshold of 750 $\mu\text{m}^2/\text{sec}$ for differentiating significant prostate cancer is suggested in Meyer's conclusion. Multiple factors can affect measured ADC values, including the b-values used[119, 120], field strength, echo times, patient and coil geometry, and interreader variability[116] (Bengtsson). Interestingly, ADC values were found to be robust across different MRI systems in a recently published phantom study[121].

ADC ratios can be used to avoid potential sources of error due to variability in absolute ADC values by relating them to ADC values in normal tissue and thereby normalizing ADC[122, 123]. This approach has been used in other organ systems[117, 124-127]. Different normal references have been used[111]. However, there are also sources of error when using internal reference organs[128].

Different definitions of the term "significant cancer" have been used in previous studies, leading to the use of ISUP cutoffs. Some have defined ISUP grade 1 as non-significant and classified ISUP 2 and higher as significant[129-131]. On the other hand, some have included ISUP 2 in the non-significant group. Several studies have evaluated the use of ADC ratios to distinguish between different ISUP grades[110, 120, 122, 132]. Regardless of the definition of clinically significant prostate cancer

used, several authors have reported a strong inverse correlation between ADC measures and tumor aggressiveness[133, 134].

In summary, DWI and ADC values offer an opportunity to quantify tumor biological properties and aggressiveness, and it is desirable to move away from qualitative subjective assessment tools. In addition, AI-models rely on the calculation of high b-values for the detection and characterization of suspected cancer lesions. Several important studies have investigated the relationship between tumor characteristics and ADC values/ADC ratios but few studies have a multicenter, approach that resembles a true clinical context. In Paper II of this dissertation, we examined this relationship in a diverse, multicenter MRI dataset with tumor maps from prostatectomies as the reference standard.

Artificial intelligence in prostate MRI

Introduction

In computer science, artificial intelligence, AI, is a collective term. A way to visualize the mutual relationship and hierarchy of different fields of AI is presented in Figure 6.

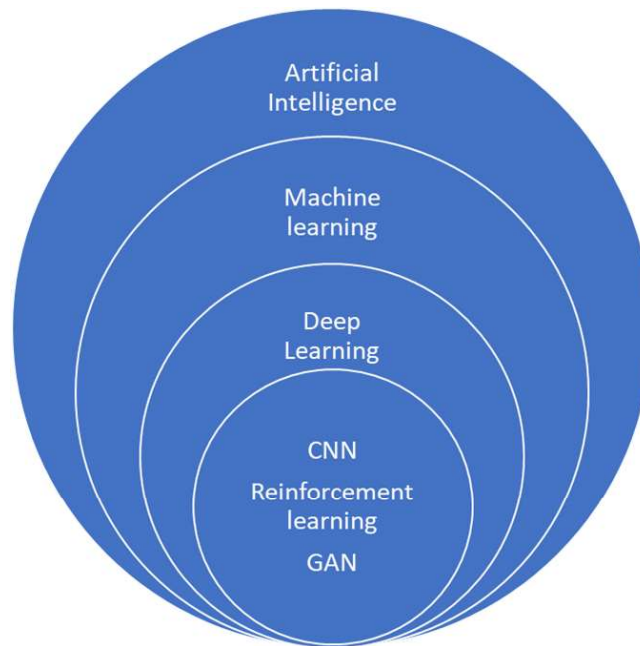


Figure 6. Illustration of the infrastructure and interrelation of AI-methods.

The development of deep neural networks has advanced rapidly in recent years, explained by a combination of improved algorithms, increased availability of training data, and enhanced computational capacity through parallel computing using graphics processing unit (GPU) technology. The AI-models used in radiology often combine various types of deep neural networks. Convolutional neural networks (CNN) utilize filters/kernels to scan training data and identify features

important for image recognition. Reinforcement learning employs rewards and punishments to guide the algorithm toward desired outcomes. Moreover, the Generative Adversarial Network model consists of two parallel networks, one generating new data that is difficult to distinguish from authentic data, while the other discriminates between genuine training data and generated data. Figure 7a outlines the schematic presentation of the structure of one of the available AI-models in prostate MRI, and Figure 7b shows an output example from this AI-model.

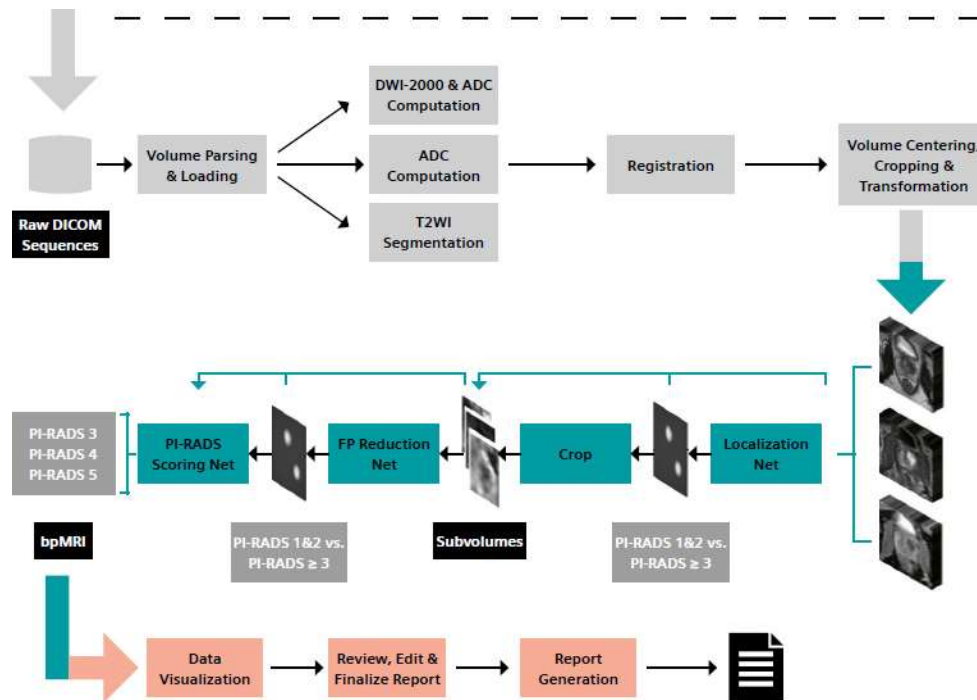


Figure 7a. Deep learning architecture with the pre-processing pipeline (gray); deep learning-based lesion detection and classification component (green). DICOM = Digital Imaging and Communications in Medicine, ADC = apparent diffusion coefficient, FP = false positive, PI-RADS = Prostate Imaging Reporting- and Data System. Adapted from Winkel et al[135].

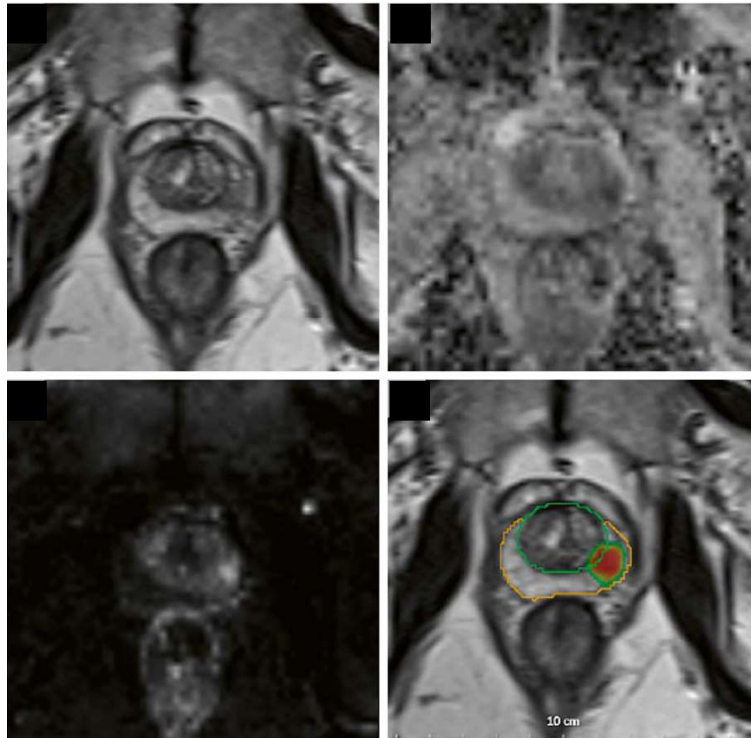


Figure 7b. A lesion in the left apical PZpl of a 51-year-old man, with a maximum diameter of 10.2 mm and a mean ADC-value of $961 \mu\text{m}^2/\text{s}$. Prostate AI detected the lesion and assigned a PI-RADS 4 category. Biopsy results revealed a Gleason $3+3 = 6$ pattern. Data visualization platform with the T2w images, ADC map, and high b-value image as well as the T2w image overlaid with the AI-generated heatmap in red. Whole gland segmentation in orange and transition zone segmentation in green. Adapted from Winkel et al [135].

AI-models in Radiology and Prostate MRI

The interest in and development of AI-models in radiology have increased exponentially in the past decade[136].

One challenge is that there is a mismatch between the pace at which new algorithms are approved and the scientific evidence[137] for their performance and, more importantly, the outcomes of their usage in clinical reality[138]. This knowledge gap is the focus for Papers I and II but my thesis does not delve into the detailed functioning of the AI-models used in prostate MRI.

AI-models have the potential to facilitate workflow in several steps of prostate MRI, and numerous published review articles highlight this topic[139-144]. In addition in 2021 a position Paper by ESUR(European Society of Urogenital Radiology)/ESUI (European Society of Urological Imaging) was published[145]. Among the tasks

expected to be facilitated by AI-models in the workflow of prostate MRI are accurate patient positioning, optimized scanning planes, and individualized sequence optimization. However, I choose to focus on three areas: image quality, prostate gland segmentation, and the detection/characterization of suspicious focal cancer lesions.

Image quality

AI-models can improve image quality, and in the past year, we have seen clinical breakthroughs for these applications[146-150]. By applying advanced algorithms, they have successfully reduced the signal-to-noise ratio and various artifacts in images for key prostate MRI sequences, turbo spin echo T2W, and DWI. This has already had a clinical impact, reducing examination times and allowing for increased patient throughput in the MRI scanners. In a possible future screening scenario, these AI-facilitated improvements have great potential to help us manage the expected large volume of examinations.

Prostate gland segmentation

Another area where about ten AI-models are commercially available[139-141, 145] is prostate segmentation. From this segmentation, we obtain prostate volume for PSAD calculation and contours that can be exported for use in fusion biopsy preparation and various radiation therapy techniques. Figure 8 shows an example from Paper II, with segmentation using two commercially available AI algorithms and a reference standard (expert radiologist).

As discussed earlier in the chapter on prostate volume, manual segmentation is time-consuming, and the dominant methods (ellipsoid formula used in TRUS and MRI) are user-dependent and come with various sources of error. To encourage development in this area, several so-called grand challenges have been conducted, where MRI data is made available for the development and training of participating teams. The actual competition involves evaluating the participants' algorithms on a test set of MRI data that has not been available previously. The most well-known challenges for prostate segmentation are Promise12[151], NCI-ISBI 2013[152], and MSD[153]. Early experimental studies in the area of segmentation with neuronal networks date back to 2013[87, 94, 154], and over the past decade, there have been numerous studies evaluating deep learning algorithms for prostate segmentation[155-157]. Although the results have been good in terms of agreement in prostate volume and segmentation compared to the different reference standards used, these studies have limitations. Typically they are conducted at a single institution, use only one camera model and field strength, require pre-processing of MRI data, have relatively limited sizes of their test sets, and, perhaps most importantly, lack validation against external cohorts[93, 94, 158-162].

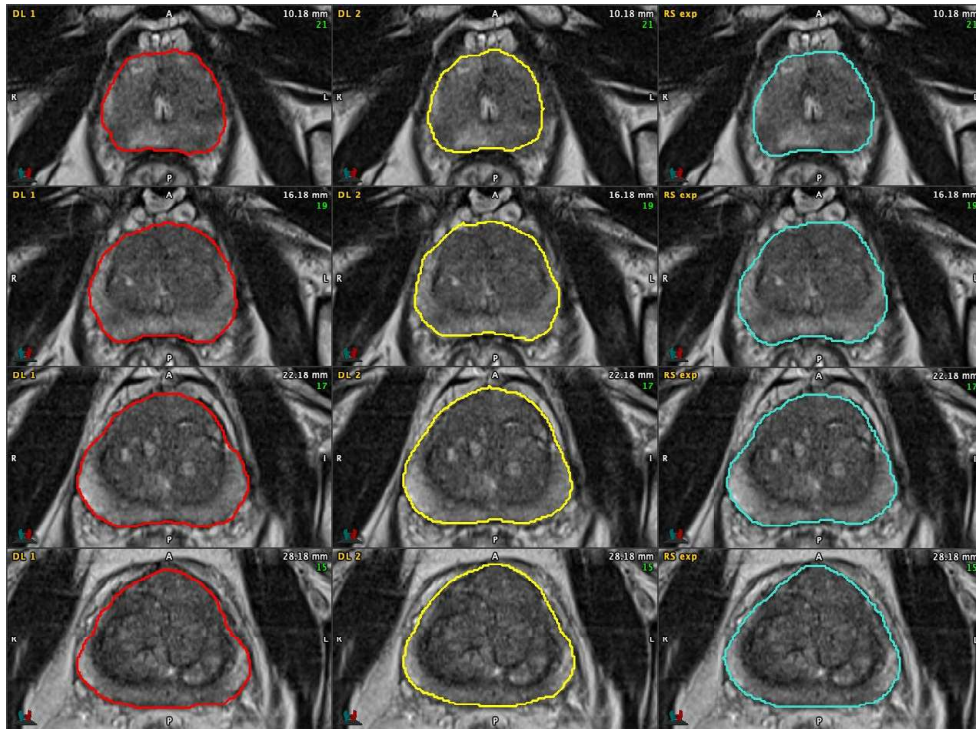


Fig. 8 Consecutive T2 transaxials with prostate contours by planimetry. Left column: Deep Learning Algorithm 1 (red contour); middle column: Deep Learning Algorithm 2 (yellow contour); right column: the reference standard by expert radiologist (turquoise contour). Both deep learning algorithms succeeded in differentiating the prostate pseudocapsule from the endopelvic fascia and extraprostatic fat (white arrows in the middle column). Failure to do this is one of the known pitfalls when performing planimetry. Adapted from Thimansson et al, Paper II.

Figure 9 shows the process chain in a case from our clinical practice with AI-assisted preparation for fusion biopsy.

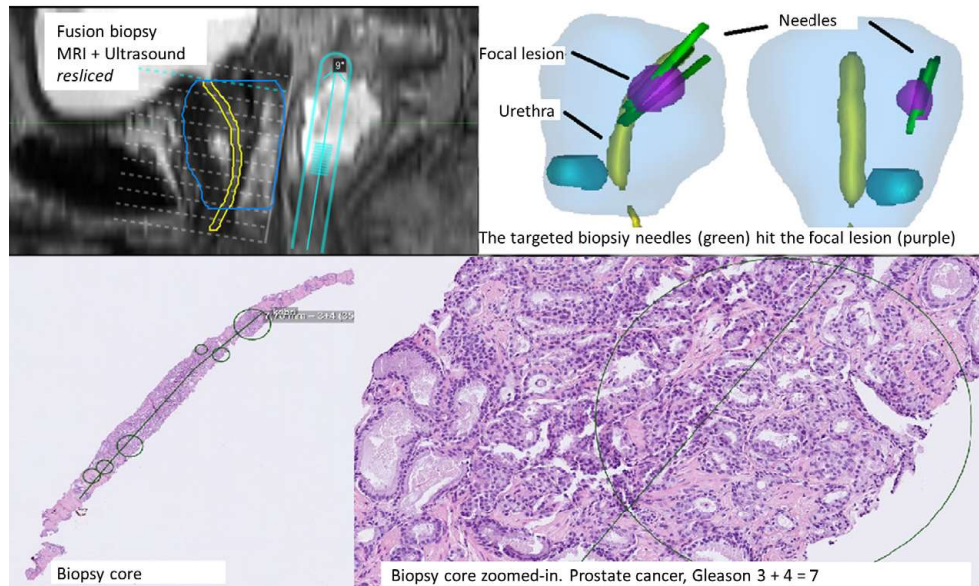


Figure 9. AI-assisted pre-processing for fusion biopsy. The top left shows segmentation (blue) of the prostate where the AI-model presents a proposal to the radiologist which is reviewed and adjusted. Urethra (yellow) and focal lesions (not shown) are also segmented. The top right shows the three-dimensional presentation of biopsy needle positions and the focal lesions (purple and blue). The bottom line shows the digitized pathology with a zoomed-in portion with Gleason 3+4=7. Collage Dr Thimansson.

Lesion detection and characterization

The next major and important area for AI-models to master in prostate MRI is the detection and characterization of focal lesions. There are four commercially available platforms[143, 144], three of which are approved for use in the European Union. However, the number is increasing, and several AI-models have recently obtained or are about to obtain approval in the EU. Studies show that deep learning-based AI-models can detect suspicious cancer lesions with high sensitivity and acceptable trade-off in terms of sensitivity and specificity, measured as the area-under-curve (AUC)[163-169]. The use of AI-models can improve radiologists' performance and efficiency in interpreting prostate MRI[170]. AI-models can be used to automate PI-RADS classification, reduce interreader variability, and studies have shown that AI-models can achieve similar PI-RADS results as experts and are more consistent in predicting high-risk lesions. The challenges lie in obtaining representative training data and demonstrating generalizability in our clinical

reality. Developing robust AI-models requires large and diverse training and testing datasets, and the availability of diverse and well-composed data is a limitation in the field. Constructing large and versatile datasets for AI-models can be challenging due to geographical differences in data protection regulations. Here, alternative solutions such as federated learning¹[171] can facilitate the development of robust AI-models across different institutions. Another challenge is that current performance measures are mainly based on cross-validation and lack real interaction between radiologists and AI systems, “these studies are far from representative of a real-world setting” according to a recent review article[143].

Several ongoing studies are tackling this challenge, and I have had the privilege of participating in the largest comparison between AI-models and radiologists, PI-CAI 22[172]. In this study, each radiologist reviewed blocks of 100 prostate MRI’s and, in addition to detecting and characterizing suspicious focal cancer lesions, also indicated the total risk for significant cancer between 1-100, a metric that AI-models provide as an output. The first results from the study were presented recently (see chapter on Future perspectives).

The OPT research consortium is currently conducting several studies in this area, comparing the performance of local radiologists, expert radiologists, and different AI-models. Our preliminary results from a pilot study (in manuscript) show that the AI-model is good at detecting significant cancer but would also select significantly more men for biopsy than radiologists do.

The challenge with the reference standards, what is really *the truth*?

The AI-models we use in prostate MRI are typically based on deep neural networks, specifically convolutional neural networks (CNNs), which consist of an input layer, hidden layers, and an output layer. These models are typically trained by feeding large amounts of training data² into the input layer. We can influence the process in the hidden layer through various techniques, but we have limited insight into the process that leads to the information presented in the output layer. However, we want to relate this information to what we use as the truth, our reference standard,

¹ Federated learning is a technique used by artificial intelligence (AI) that enables the training of models on data distributed across multiple institutions without centralizing the data collection. Instead of transmitting data to a central server for training, the model's code is sent to the devices, which then locally train the model on their own data. Subsequently, only the updates from each device are sent back to the server for aggregation and model updating.

² With varying quality and levels of annotation - the better the quality of the training data, the greater the potential for high-quality output.

and therefore we use a so-called supervised learning approach, where we have labeled our dataset in terms of the desired output. Hence, we have good opportunities to fine-tune our AI-model by rewarding it when it produces output that aligns with our desires and punishing it for incorrect output.

The challenge that remains is what we should use as the truth to evaluate whether AI-models have performed well or poorly. In the context of prostate MRI, applicable examples are: what is the true prostate volume and what are the true cancer-suspected lesions? If we rely on expert consensus from radiologists, there will be inter-observer variability as experts may have somewhat, but not entirely, similar opinions. If we use histology from biopsies or tumor maps as the truth for what constitutes a cancer-suspected lesion, uncertainties regarding the accuracy of biopsies and differences in assessments by pathologists arise.

Therefore, to conclude, a limiting factor in the current methodology for the development and evaluation of AI-models is the uncertainty in our reference standards, which is well summarized in this quote I received in a personal mail correspondence with an authority in AI-models for prostate segmentation, Professor Henkjan Huisman: *“AI algorithms can easily reproduce the reference standards when looking at expert variability. Showing a better algorithm requires creating a better reference standard. This is a hard problem for which I have no solution. Experts disagree too much, so that does not help. Let me know if you think of a solution.”*

Organized Prostate Cancer Testing (OPT)

Screening

Four criteria should be met according to the World Health Organization (WHO) before considering general screening for a health condition[173]. 1, The condition should be significant from a public health perspective, be common and/or entail significant suffering and illness for the individual. This requires nuanced considerations, taking into account factors such as prevalence, age at onset, prognosis, treatment complications, costs, and other relevant aspects, The diagnostic method should have high sensitivity and specificity, and the method must be accepted by the target population. There should be treatment that alters the natural course of the disease in a favorable direction., The costs should be in reasonable proportion to the effects, including the economic costs of the screening program (including supplementary diagnostics and treatment), as well as “soft costs” in the form of physical and psychological discomfort, which should be included in the analysis.

Cancer screening guidelines vary among nations, sharing some similarities but also exhibiting distinct differences. While well-established screening programs for breast and cervical cancer have common recommendations, there is more divergence among countries when it comes to screening for prostate, colorectal, lung, and skin cancer[174].

In Sweden and other developed nations, non-organized prostate cancer testing is prevalent and has contributed to an increase in the diagnosis of low-volume tumors over the past few decades[175-177]. However, non-organized prostate cancer testing is ineffective and shows socioeconomic disparities[178, 179].

Screening for prostate cancer using prostate-specific antigen (PSA) and systematic biopsies over an extended period may lower prostate cancer mortality but leads to unacceptable levels of overdiagnosis[6]. Utilizing risk calculators, biomarkers, and MRI to choose men with elevated PSA for biopsy and opting for targeted biopsies rather than systematic ones reduces the identification of low-grade cancer, consequently mitigating overdiagnosis[180]. A recent study highlights the importance of MRI showing that a substantial proportion of significant cancers were

found by MRI in men without increased PSA levels[181]. Nonetheless, there are substantial gaps in knowledge, such as the effectiveness of contemporary diagnostic methods in long-term screening programs and their impact on mortality. These gaps are presently under examination in three significant randomized screening studies, Gothenburg-2 trial (Sweden), ProScreen (Finland) and PROBASE (Germany).

Organized testing

The Swedish National Board of Health and Welfare has three times, most recently in 2018[182] advised against the implementation of national screening for prostate cancer. The main argument behind this recommendation is that the benefits of prostate cancer screening with PSA testing do not clearly outweigh the negative effects, and knowledge about supplementary testing methods is still insufficient. However, in 2018, it was emphasized that it is highly important for research on diagnostics to progress so that more men with severe prostate cancer can receive early medical attention.

Consequently, the Swedish Ministry of Health and Social Affairs tasked the Confederation of Regional Cancer Centers with implementing measures to establish standardized and effective prostate cancer testing in Sweden. The objective was to enhance understanding of diagnostic methods capable of selectively identifying clinically significant, early-stage prostate cancer. In response to the ministry's directive, the Confederation of Regional Cancer Centers devised regional initiatives incorporating "organized prostate cancer testing" (OPT)[9].

At the EU level, the recommendations for cancer screening were revised in 2022[8, 183]. The updated recommendation identified prostate cancer as a suitable target disease for organized screening, encouraging countries to initiate pilot programs and conduct additional research. In 2020, Västra Götalandsregionen and Region Skåne became the first two health regions in Sweden to implement organized prostate cancer testing (OPT), as detailed in a methodology study published in Eur Urol Focus 2023[88].

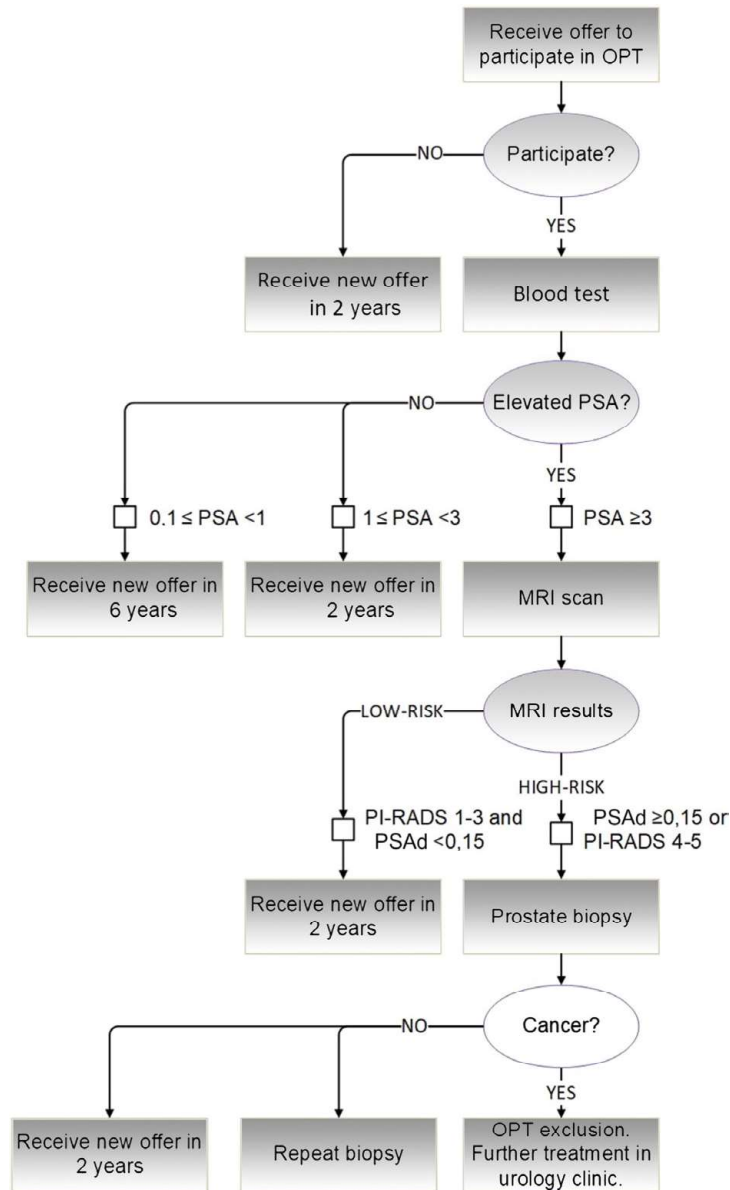


Figure 10. The original diagnostic pathway in OPT. Published with permission from Eur Urol Focus[88].

The diagnostic pathway in OPT (Figure 10) begins with an informatively balanced letter sent via the administrative system outlining potential advantages and disadvantages of undergoing the PSA test. For men with PSA levels $\geq 3 \mu\text{g/l}$, the OPT coordinator is alerted by the administrative system, prompting the issuance of a standardized referral for an MRI scan at an affiliated radiology department.

Standardized MRI reports encompass prostate volume, PSA density, and diagnostic evaluations based on the PI-RADS v2.1, indicating lesions' anatomical locations. The OPT office records the MRI report in the participant's medical record, the regional OPT register, and the national cancer registry, with automatic updates from histological assessments of biopsy cores. A comprehensive urological investigation protocol, developed by a team of prostate cancer experts, directs individuals with a suspicious MRI lesion (PI-RADS 4–5) and/or PSA density $\geq 0.15 \mu\text{g}/\text{l}/\text{cm}^3$ to an OPT-associated urology department. Those with PSA levels $\geq 3 \mu\text{g}/\text{l}$ unable to undergo MRI are directly referred to a urology department. The algorithm adheres to national prostate cancer guidelines, allowing for individualized management in specific situations. The algorithm determines whether biopsies should be systematic, targeted, or a combination of both. The OPT office records all pathology report details in the regional OPT register. Upon a prostate cancer diagnosis, the patient transitions to standard care. Those with benign biopsy results are automatically re-invited for a PSA test after two years, unless the OPT algorithm prescribes earlier diagnostic investigations for specific reasons. Figure 11 depicts an authentic case from OPT.

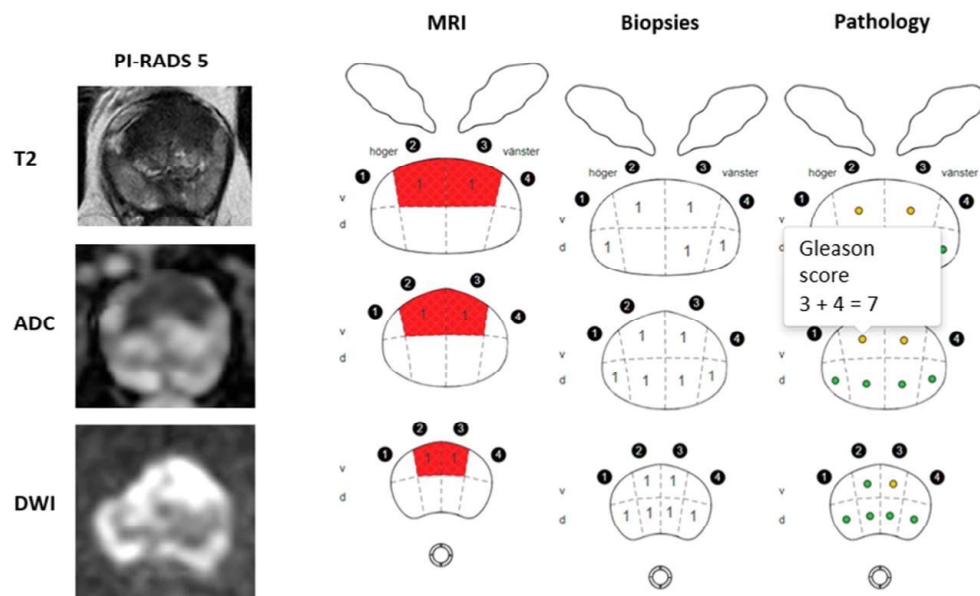


Figure 11. Results for a 50-yr-old man with prostate-specific antigen of 4.9 ng/ml and prostate-specific antigen density of 0.15 ng/ml/cm³. Left: T2-weighted transaxial magnetic resonance imaging (MRI) and apparent diffusion coefficient (ADC)/diffusion-weighted imaging (DWI) with a focal lesion in the ventral portion of the transition zone base with a Prostate Imaging-Reporting and Data System (PI-RADS) score of 5. The corresponding INCA report showing the MRI lesion (left), the number and position of biopsy cores (middle), and histology feedback from the pathologist (right). In this case targeted biopsies were positive ventrally (indicated by yellow) and negative dorsally (indicated by green) in the histology feedback template. INCA = Swedish nationwide web-based register platform for cancer patient data. Published with permission from Eur Urol Focus[88].

A short introduction to the key statistical methods used in the thesis.

Bland-Altman plots

Bland-Altman plots are widely used in statistics and research to evaluate the agreement between two measurement methods or observers. Introduced by Martin Bland and Douglas Altman in 1986[184], these plots visually represent the differences, termed "bias" or "mean difference," between two sets of measurements against their average. The creation of these plots involves calculating differences between paired measurements and plotting them against their means on a scatter plot. The mean difference is depicted as a horizontal line at zero, signifying perfect agreement, while parallel lines, known as "limits of agreement," indicate the expected range for 95% of the differences. Interpreting Bland-Altman plots involves examining patterns in differences and assessing their distribution around the mean difference. Evenly distributed differences around the zero line suggest good agreement, while systematic bias or patterns indicate disagreement. The limits of agreement are critical, representing the range where 95% of differences are expected. Typically calculated as the mean difference ± 1.96 times the standard deviation, exceeding these limits may suggest poor agreement. Bland Altman plots is the important statistical method in Paper I for method agreement and is also used in Paper II for method agreement and in Paper III for interreader variability.

DICE coefficients

DICE coefficients, also known as Sørensen-DICE coefficients[185], serve as a statistical metric for assessing the similarity between two sets or samples. This coefficient quantifies the degree of overlap or agreement between data sets. It ranges from 0 to 1, with 0 indicating no overlap and 1 representing complete similarity. In image processing, it evaluates similarity in segmented regions. A high DICE coefficient suggests strong agreement or similarity, indicating reliable results. DICE coefficients is the principal statistical model in Paper II for evaluation of deep-learning segmentation.

Spearman's rank correlation coefficient

Spearman's rank correlation coefficient[186] quantifies the strength and direction of the relationship between two, not normally distributed, variables. It does not assume any specific distribution of the data. When the relationship between the two variables is not linear it is not appropriate to use traditional correlation measures, such as Pearson's correlation coefficient. The coefficient is calculated by assigning ranks to both variables, then computing the difference between these ranks for each observation using a specific formula. The resulting coefficients range from minus 1 to plus 1, where a coefficient close to plus one or minus one indicates a strong positive or negative correlation respectively. A coefficient close to zero means that there is no correlation between the variables. Spearman's rank correlation coefficient is used in Paper III to evaluate the association between ISUP grade and ADC variables.

Receiver Operating Characteristic (ROC) curves

Receiver Operating Characteristic (ROC) curves are graphical representations used to assess the performance of binary classification models and provide insights into the trade-off between sensitivity (true positive rate) and specificity (true negative rate), aiding decision-making based on application-specific priorities. The ROC curve is constructed by plotting sensitivity against the false positive rate (1 - specificity). The false positive rate quantifies the model's tendency to wrongly predict positive occurrences, while the true positive rate reflects its accuracy in identifying positive instances. The area under the ROC curve (AUC-ROC) serves as a quantitative measure of model performance, ranging from 0 to 1. Higher AUC-ROC values indicate superior performance, with 0.5 implying no better accuracy than random guessing and 1 representing a perfect model. The ROC curve facilitates model comparison, with higher AUC-ROC values generally indicating more accurate discrimination between positive and negative instances. ROC-curves were used in Paper III to evaluate the ability to discriminate between ISUP 1-2 and ISUP 3-5 based on ADC variables.

Personal ethical reflections

It is the responsibility of every researcher, to the best of their ability, to make active ethical choices that ensure their research contributes positively to a larger context. These choices are made on a small scale, such as obtaining informed consent for study participation, but we as researchers need to adopt a helicopter perspective on our research and consider the larger scale ethical questions for our research direction.

In this context, I have two concerns that can be summarized as follows:

1 Does prostate cancer screening extend lives?

To justify general screening, the WHO criteria must be met, and the fourth criterion (see screening chapter) mandates that costs should be proportional to benefits. If we turn healthy men into patients, we must ensure that the benefits outweigh the harms. There is strong evidence that PSA-based screening reduces prostate cancer mortality but at the expense of overdiagnosis, leading to physical and psychological complications[6, 187]. With the addition of MRI, we believe we can reduce the downsides and decrease overdiagnosis while maintaining an acceptable detection of significant cancer. Timely diagnosis and timely treatment are other important concepts with ethical implications; the impact of giving a cancer diagnosis to a previously healthy 50-year-old man is likely to affect quality of life[188], but could we wait 10-15 years to diagnose without real consequences given the generally slow growth of prostate cancer? How does one's well-being change from living many years with a cancer diagnosis while waiting for potential treatment, e.g., in the scenario of active surveillance?

It would be easier for me as a researcher to ethically navigate these questions if we knew that screening actually extends life and did not just reduce *cancer specific mortality*, but we do not know that today. Even the long-established breast cancer screening cannot definitively answer that important question[189].

2. Is Artificial Intelligence a help or hindrance?

Since the Dartmouth meeting in 1956[190], research and development in the field of AI have been ongoing with varying intensity, through several so-called AI

winters, when belief in the concept weakened and resources were lacking, to the rapid development over the past five years when the availability of resources such as data, venture capital, and graphics card-based (GPU) computing power has propelled AI development forward by leaps and bounds. In the research domains I have been involved in we deal with supervised learning, where we may not know what happens inside the black box, but we determine if the output is correct in relation to our benchmark. This contrasts with the concept of AGI, Artificial General Intelligence, where we face an entity with abilities surpassing human capabilities across the broad spectrum of competencies associated with intelligence, and as a logical consequence, it can continuously improve itself[191]. Leading AI researchers debate when a potential AGI breakthrough is on the horizon[192, 193] and how humanity can ensure that AGI serves us well, for example, by curing all cancer and solving the climate crisis, instead of heralding the end of our civilization, exemplified by the Paperclip Armageddon[190, 194].

Ethically, it would be easier for me as a researcher if I knew that artificial intelligence is beneficial for humanity because, in some way, I contribute to and legitimize the use of artificial intelligence through my work.

Summary of the studies

Ethical approvals

The local ethics review committee at Lund University approved the studies in Papers I-III (entry no. 2014-886). The Swedish Ethical Review Authority approved an addendum for Papers I-III (entry no. 2019-03674) and also approved the study in Paper IV (2020-03923 and 2021-06647-02).

Cohorts

The study cohorts for Papers I-III are based on a common population consisting of consecutive patients who underwent radical robot-assisted prostatectomy at Skåne University Hospital in 2018. Shared exclusion criteria for the studies encompassed patients who had not undergone preoperative MRI of the prostate. Based on the remaining 144 patients, a number of participants were excluded specifically from each study, which is visualized in the flow chart in Figure 12.

For Paper IV a representative cohort of 999 men, aged 50 ($n = 367$), 56 ($n = 327$), and 62 years ($n = 305$), was created by randomly selecting individuals from 33 municipalities in Region Skåne, using the Swedish population registry.

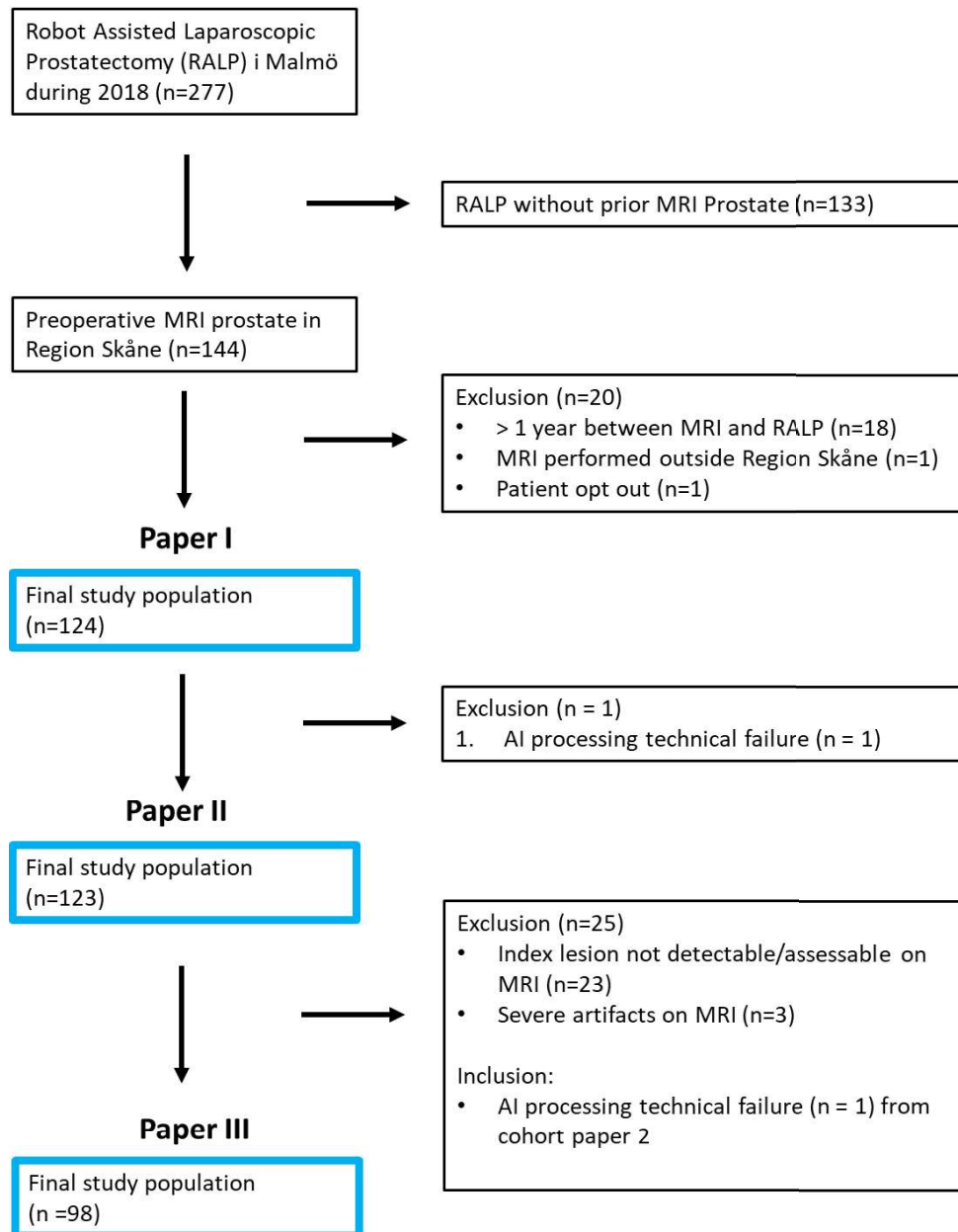


Figure 12. overview of cohorts for Papers I-III.

MRI Dataset

In our effort to reflect a real clinical setting, the prostate MRI examinations in Papers I-III originate from different hospitals and were conducted with diverse protocols and parameters. The various exclusion criteria outlined in Figure 12 explain the variations among the MRI datasets, with the most significant difference numerically observed in Paper III compared to Papers I and II. In Papers I and II, the T2W sequence is the most important for both manual prostate volume measurement, involving whole gland segmentation by the radiologist, and for the AI-models. In Paper III, the crucial sequence is the DWI. Various platforms (scanner models and magnetic field strengths) and parameters (variations in MRI protocols) are elaborated upon in the corresponding study.

Methods

The study presented in **Paper I** was a retrospective multicenter investigation focusing on diverse approaches to compute prostate volume. The study involved 124 consecutive patients who had undergone MRI prostate within a year before robot assisted radical prostatectomy surgery at Malmö University Hospital in 2018. MRI scans were conducted across seven hospitals using eight scanners, encompassing seven scanner models and two field strengths (1.5 and 3 Tesla). Different imaging acquisition parameters were employed based on local protocols. Eight distinct prostate volumes were determined through various methods (see Figure 13). Two radiologists independently calculated prostate volumes using the ellipsoid formula method (a in Figure 13). Manual planimetry, performed by two radiologists, involved tracing prostate boundaries on T2 images in three planes using external software (c in Figure 13). One highly experienced radiologist executed this method, while another with less experience performed it on a different server to maintain blinding. The AI-model was a deep learning algorithm utilizing a convolutional neural network for whole gland segmentation on non-annotated T2 images (d in Figure 13). Clinical routine (TRUS) prostate volumes, obtained during urology department visits, were collected (b in Figure 13). Prostate volume from surgical specimens was determined using dimensions in three planes and weight from pathology reports, employing the ellipsoid formula and prostate density coefficient respectively (e in Figure 13). The statistical analysis incorporated descriptive statistics, box plots, and Bland-Altman plots. The main focus was to provide a method agreement comparison between today's gold standard (radiologist-dependent ellipsoid formula) and a commercially available AI-model using two different reference standards, manual planimetry by expert radiologist, and specimen weight volume. Interreader agreement was analyzed between an

experienced and an inexperienced radiologist performing manual planimetry, as well as between two radiologists performing ellipsoid formula volume assessment.

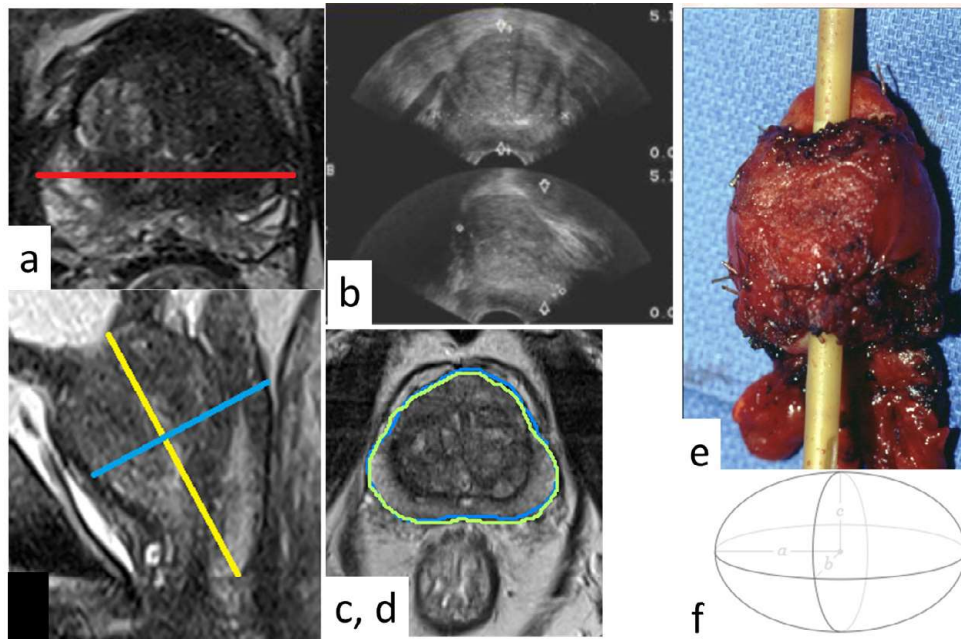


Figure 13. Overview of the different prostate volume assessment in Paper 1[91]. a, Ellipsoid formula on T2W. b, TRUS. c, planimetry by radiologist (blue) d, planimetry by the AI-model (green). e, specimen from prostatectomy. f, the formula for volume of an ellipsoid (used in a, b and e).

The multicenter study presented in **Paper II** aimed to retrospectively evaluate the performance of whole gland segmentation by two commercially available AI-models. The MRI examinations were performed across seven distinct hospitals, utilizing eight scanners with varying models, field strengths, and slice angulations. A consultant radiologist who was highly experienced in the fusion biopsy pathway executed prostate gland segmentation through manual planimetry with external software. This time-consuming process entailed fitting a line to the outer contour of the prostate on all T2 transaxial slices, referencing the coronal and sagittal slices. These entire gland segmentation contours served as reference standards in the investigation. T2W transversal images were manually transferred from the Picture Archive and Communications System to the AI-models for segmentation. The resulting contours were then exported to a server for comparison with the reference standard. The authors gauged the segmentation overlap between the deep learning algorithms and the reference standard using the DICE similarity coefficient and the Jaccard index. Evaluation of segmentation accuracy employed the calculation of the

Hausdorff distance. Descriptive statistics, including mean, standard deviation, median, interquartile range, and minimum-maximum values, were presented for the evaluation metrics. A paired t-test was used to compare DICE coefficients for the two AI-models. Bland-Altman plots were used for prostate volume method agreement between the two AI-models in comparison to the expert radiologist's volumes from segmentation.

In **Paper III** we conducted a retrospective cohort study to explore the correlation between ADC metrics and ISUP grade in patients undergoing robot-assisted laparoscopic prostatectomy (RALP) for biopsy-confirmed prostate cancer (PCa). Prostate MRIs were performed preoperatively with varied imaging parameters at different sites within Region Skåne. Sequences included transverse, coronal, and sagittal T2-weighted images, transverse T1-weighted images, diffusion-weighted images with a high b-value, and a calculated ADC map. Two radiology specialists performed imaging analyses, matching the surgical specimen's index lesion with the corresponding MRI lesion, recording the maximum diameter, zone location, and PI-RADS score. Region of interest (ROI) measurements on the ADC map covered various regions, including the index lesion, the contralateral position, the most homogeneous peripheral zone area, and the urinary bladder. Descriptive statistics presented the study population, while box plots and Spearman's rank correlation coefficient assessed the ISUP grade's association with ADC variables. Receiver operating characteristic (ROC) curves gauged discriminatory ability among different ISUP grades based on ADC metrics. Bland-Altman plots and intraclass correlation (ICC) evaluated the interreader variability. Figure 14 shows the methodology used for the different ADC-metrics used in Paper III with a corresponding histological tumor map.

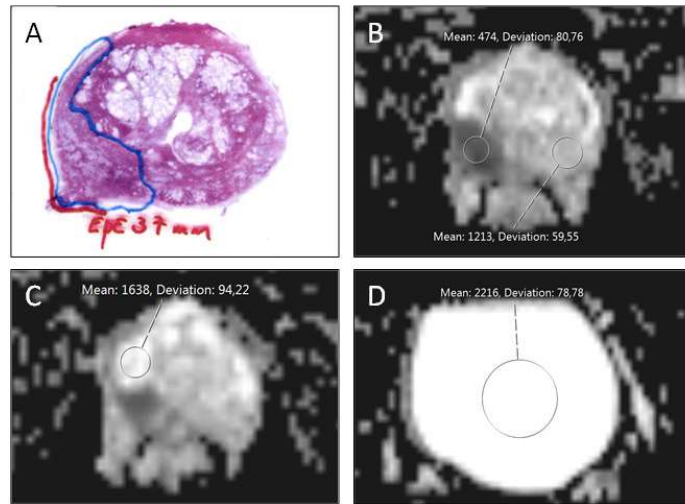


Figure 14. Tumor map from a pathology specimen and placement of ROI's in the ADC map from Paper III. A Tumor map with a large tumor in the right peripheral zone (blue border) with a 37 mm extraprostatic extension (red line). B, circular ROI in the tumor and in contralateral non-tumorous tissue. C, Circular ROI drawn in non-tumorous PZ and D, circular ROI in the urinary bladder.[116]

Paper IV presents a pilot study that assessed the feasibility of a digitally automated population-based program for organized prostate cancer testing in Region Skåne. Invitations with detailed information, PSA referrals, and consent forms were sent to potential participants. PSA testing and analysis were carried out at regional laboratories, with the results automatically recorded. The management algorithm involved categorizing PSA levels, with those below 3 $\mu\text{g/l}$ reassigned to the digital watchlist. PSA levels above 100 $\mu\text{g/l}$ prompted immediate urological assessment, while levels between 3 and 99 $\mu\text{g/l}$ led to a prostate MRI based on PI-RADS 2.1 guidelines. Further assessments, including DRE and TRUS, were performed for PSA levels of 3 $\mu\text{g/l}$ or higher, guiding prostate biopsies. Biopsy results were histologically evaluated according to ISUP 2014 consensus and ISUP Gleason Grade group (GG) classification. Patients with positive biopsies were transferred for urological evaluation and treatment, while those with negative results returned to the digital watch list for a new invitation after two years, except for those meeting specific criteria. Descriptive statistics analysed various data points, including municipal registry information, invitation numbers, participation rates, lead times, PSA levels, MRI outcomes, PSAD values, prostate biopsy cancer detection rates, and treatment choices.

Results

The study in **Paper I** included 124 patients out of the 144 patients who underwent RALP (exclusion criteria in fig 12), with a median age of 66 years (range: 45-76) and a median preoperative PSA of 6.90 $\mu\text{g/l}$ (min: 0.88, max: 39). As shown in Figure 15a, the mean difference between prostate volumes from the AI-model (PVDL) and the reference standard manual planimetry by expert radiologist (PVMPE) was lower than between two radiologists using the ellipsoid formula (PVEF1 and PVEF2) and the reference standard (PVMPE) (mean difference and limits of agreement PVDL: -0.33 mL [-10.80; 10.14 mL], PVEF1: -3.83 mL [-19.55; 11.89 mL], PVEF2: -3.05 mL [-18.55; 12.45 mL]). PVDL showed slightly narrower limits of agreement than PVEF, indicating better precision. In a sensitivity analysis using specimen weight prostate volume (PVSW) as the reference standard (Figure 15 b), the mean difference and limits of agreement were lower for PVDL than for PVEF1 or PVEF2, confirming higher accuracy for PVDL (mean difference PVDL: -4.22 mL [-22.52; 14.07 mL], PVEF1: -7.89 mL [-30.50; 14.73 mL], PVEF2: -6.97 mL [-30.13; 16.18 mL]). Both Bland Altman plots in Figure 15 showed a tendency of the AI-model and the ellipsoid formula to underestimate volumes of enlarged prostates compared to expert manual planimetry and specimen weight, with the AI-model showing a lesser extent of underestimation than the ellipsoid formula. The interreader agreement in manual planimetry between unexperienced (PVMPU) and experienced reader (PVMPE) showed that the unexperienced reader calculated lower volumes than the experienced reader but with better precision than between the two radiologists using the ellipsoid formula: mean difference and limits of agreement PVMPE vs. PVMPU: -3.73 mL [-11.90; 4.45 mL], $p < 0.001$) and between PVEF1 and PVEF2: -0.78 mL (-15.08; 13.51 mL). For PVEF1 and PVEF 2 the intra class correlation ICC with 95% CI was 0.93 (0.96; 0.953). Summarized mean, median, minimum, and maximum values for all compared volumes were reported. Timed observations for planimetry by an experienced reader in 14/124 patients showed a mean time consumption of 8 min 4 sec per case.

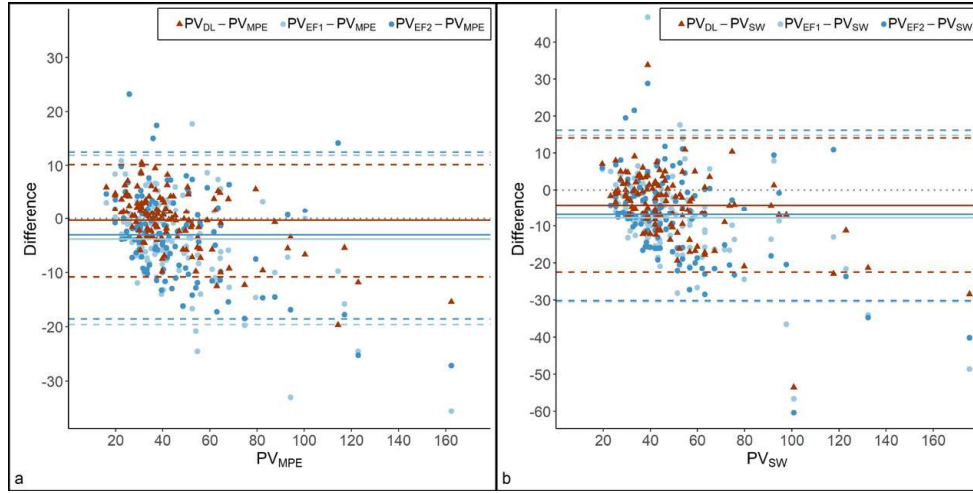


Figure 15. Bland Altman plots were employed to compare two prostate volume measurement methods: automated deep learning (PVDL) and manual ellipsoid formula performed by two radiologists (PVEF1 and PVEF2). In Figure a, prostate volumes from manual planimetry by an experienced radiologist (PVMPE) served as the reference standard, while in Figure b, prostate volumes from prostatectomy specimen weight (PVSW) were used as the reference standard. The solid lines in both plots represent the mean difference between methods, and the dashed lines depict limits of agreement, calculated as the mean difference \pm 1.96SD. The dotted line signifies no difference between methods. From Paper I[91].

The retrospective multicenter study encompassing 123 patients in the final study cohort in **Paper II** focused on comparing the performance of two certified commercially available AI-models (DL1 and DL2) for prostate gland segmentation. Concerning the DICE similarity coefficient, the mean and standard deviation for DLA1 versus the radiologist reference standard (RSexp) was 0.90 ± 0.05 , and for DLA2 versus RSexp, it stood at 0.89 ± 0.04 . The median and inter quartile range IQR/min-max for DLA1 were 0.91 (IQR 0.04, min-max 0.60–0.95), and for DLA2, it was 0.91 (IQR 0.04, min-max 0.73–0.94). The mean Jaccard index remained consistent for both DLA1 and DLA2 at 0.81. Regarding the Hausdorff distance, DLA1 recorded a slightly lower value of 7.1 mm compared to DLA2's 7.9 mm. A detailed breakdown of statistics is available in Table 2. A paired t-test comparing DICE Similarity Coefficients for DLA1 and DLA2 indicated no statistically significant difference ($p = 0.8$). In the Bland-Altman plot (Fig. 4), Using Bland Altman plots the observed mean difference in prostate volume between DLA2 and RSexp was less than the observed mean difference between DLA1 and RSexp (mean difference [95% limits of agreement]): DLA1, -3.53 mL (-11.55 ; 4.50 mL); DLA2, 0.67 mL (-6.41 ; 7.75 mL). DLA2 showed superior precision with narrower limits of agreement. Notably, DLA2 tended to overestimate the volumes of small- and medium-sized prostates, while DLA1 tended to underestimate the volumes.

Table 2. DICE similarity coefficient, Jaccard index, and Hausdorff distance mean and SD and median and IQR/min-max for the reference standard (RSexp) and deep learning algorithms (DLA1 and DLA2).

Name	Mean	SD	Median	IQR	Min	Q1	Q3	Max
DICE RFexp vs. DLA1	0.895	0.0473	0.905	0.0355	0.595	0.884	0.920	0.948
DICE RFexp vs. DLA2	0.894	0.0386	0.905	0.0415	0.726	0.879	0.920	0.941
Jaccard RFexp vs. DLA1	0.813	0.0694	0.826	0.0595	0.424	0.792	0.851	0.901
Jaccard RFexp vs. DLA2	0.811	0.0599	0.827	0.0680	0.569	0.784	0.852	0.889
Hausdorff RFexp vs. DLA1	7.11	2.80	6.16	3.80	3.28	4.95	8.75	16.6
Hausdorff RFexp vs. DLA2	7.93	3.44	7.21	3.04	3.59	5.97	9.01	23.6

Paper III involved 144 men who underwent RALP due to PC and also underwent preoperative prostate MRI. The final analysis included 98 patients after exclusion for various reasons as depicted in Figure 12. The majority of index lesions were situated in the peripheral zone of the prostate, and none of the specimens were categorized as ISUP 1. ADC measurements were conducted (visualized in Figure 14) revealing an average ADC in the (index) lesion of $652 \times 10^{-6} \text{ mm}^2/\text{s}$ (range $396 \times 10^{-6} \text{ mm}^2/\text{s}$ to $1271 \times 10^{-6} \text{ mm}^2/\text{s}$). The average ADC value in the contralateral ref tissue was $1275 \times 10^{-6} \text{ mm}^2/\text{s}$ (range $779 \times 10^{-6} \text{ mm}^2/\text{s}$ to $1794 \times 10^{-6} \text{ mm}^2/\text{s}$). The average ADC-value in the normal peripheral zone was $1478 \times 10^{-6} \text{ mm}^2/\text{s}$ (range $779 \times 10^{-6} \text{ mm}^2/\text{s}$ to $2155 \times 10^{-6} \text{ mm}^2/\text{s}$), and the ADC-value in urine ref was $2021 \times 10^{-6} \text{ mm}^2/\text{s}$ (range $861 \times 10^{-6} \text{ mm}^2/\text{s}$ to $3368 \times 10^{-6} \text{ mm}^2/\text{s}$) (see Figure 16). No significant negative correlation emerged between the ADC-value of the index lesion and ISUP grade. The Spearman correlation between the ADC-value of the index lesion and the ISUP grade was low ($r = -0.18$) and not statistically significant. The ADC-value of the index lesion failed to effectively differentiate between ISUP 1-2 and ISUP 3-5 (AUC= 0.62 [95% CI 0.51-0.74]). While a negative correlation tendency was noted in the analysis of 3T scanners ($r = -0.27$; $p < 0.05$), no such trend was observed for 1.5T scanners ($r = -0.01$). Separate analyses of the peripheral zone and transition zone yielded no correlation. The three distinct ratios related to tumor aggressiveness, namely ADC lesion/ADC contralateral reference, ADC lesion / ADC urine, and ADC lesion/ADC peripheral zone reference, did not show any capacity for discrimination. Reader agreement for ADC measurements was high: ADC lesion (ICC 0.80 [95% CI 0.72 – 0.86]), ADC contralateral reference (ICC 0.82 [95% CI 0.75 – 0.88]), ADC urine (ICC 0.96 [95% CI 0.94 – 0.97]), and ADC normal peripheral zone ref (ICC 0.75 [95% CI 0.65 – 0.86]).

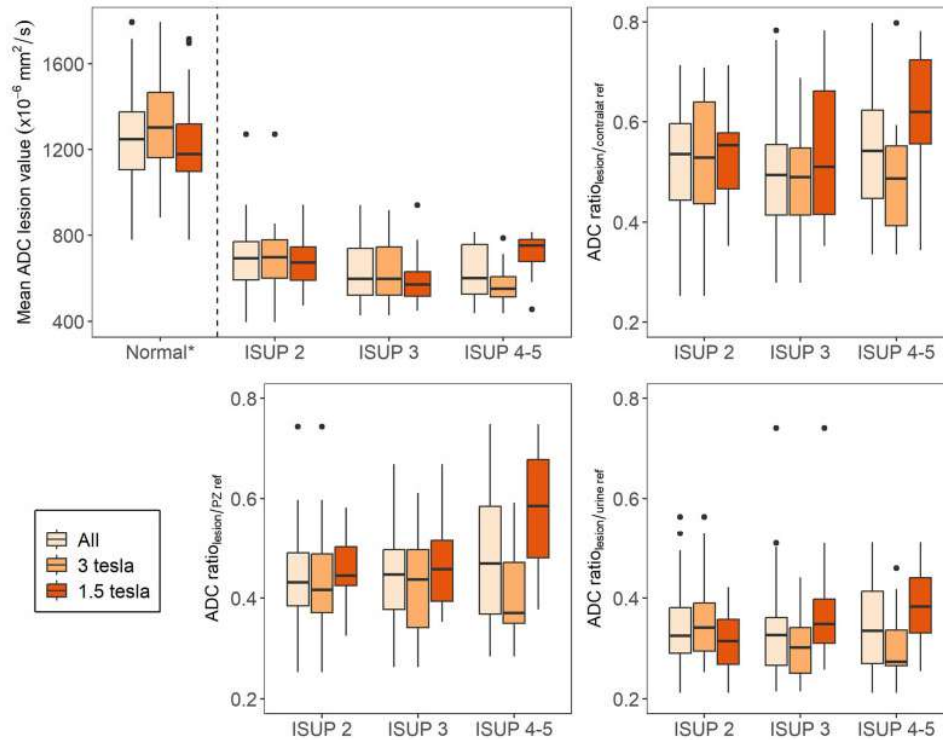


Figure 16. Box-and-whisker plots of apparent diffusion coefficient (ADC) metrics for tumors stratified by ISUP grade. (*) Normal represents the absolute ADC value of the normal-appearing tissue in the contralateral position of the index lesion.[116]

In **Paper IV** 418 (42%) out of the 999 invited individuals took part and underwent a PSA test. The participation rates for individuals aged 50, 56, and 62 years were 139/367 (38%), 143/327 (44%), and 136/305 (45%), respectively. The digital management system and OPT head office in Region Skåne handled all 999 invitations, data retrieval (PSA, MRI, and biopsy results) from the 418 participants, adherence to the algorithm, and the management of all response letters and future invitations. Among the 418 participants, 35 (8%, 95% CI 6%–11%) had PSA levels $\geq 3 \mu\text{g/l}$. The distribution across age groups was one of 139 individuals aged 50 years (0.7%, 95% CI 0.02%–4%), 10 of 143 individuals aged 56 years (7%, 95% CI 3%–12%), and 24 of 136 individuals aged 65 years (18%, 95% CI 12%–25%). Those with elevated PSA levels had a median (IQR) PSA level of 4 (3.2–5.8) $\mu\text{g/l}$, and those diagnosed with PC had a median (IQR) PSA level of 5.6 (3.43–36.5) $\mu\text{g/l}$. The number of individuals with low (1–2.9 ng/mL) and very low ($< 1 \text{ ng/mL}$) PSA levels decreased in older age groups. Among 33 individuals with elevated PSA levels who had MRI (Table 3), 16 (48%, 95% CI 31%–66%) had a negative MRI (PI-RADS ≤ 2), and 17 (52%, 95% CI 34%–69%) had a positive MRI. Of those with positive results, seven were PI-RADS 3 (21%, 95% CI 9%–39%), nine were PI-

RADS 4 (27%, 95% CI 13%–46%), and one was PI-RADS 5. The median (IQR) time from PSA test to MRI report was 24 (13.5–29) days, from MRI report to urological visit 16 (12–20) days and the time between invitation and a visit to a urological center was 65 (52–76) days. Sixteen individuals underwent prostate biopsies, with 13 of 33 individuals who had MRI scans biopsied based on PSAD values $> 0.15 \mu\text{g/l /cm}^3$ and/or PI-RADS > 3 . Systematic, targeted, and combined biopsies were conducted, with mean biopsy numbers of 10.8, 4.5, and 12.6 per individual, respectively. Clinically significant PC was found in seven individuals (1.7%), while three (0.7%) were diagnosed with clinically insignificant PC (ISUP GG < 2). The overall PC detection rates were 1% (10/999) among those invited and 2.4% (10/418) among those who participated in PSA testing. Management strategies for diagnosed PC cases included active surveillance (four individuals), robot-assisted radical prostatectomy (three), external beam radiation therapy (two), and androgen-deprivation therapy (one). All three individuals who underwent RALP had significant tumors corresponding to MRI lesions and biopsy findings. Two individuals on active surveillance underwent RALP within two years, revealing GG2 and GG3 tumors corresponding to initial PI-RADS 4 lesions in both cases.

Table 3. Distribution of men with elevated PSA level ($\mu\text{g/l}$), PSAD ($\mu\text{g/l/cm}^3$), the number of men who underwent biopsies, and the cancer-specific outcomes for those in each PI-RADS group and for those who did not undergo MRI. Data from Paper IV[89].

	MRI not done	PI-RADS<3	PI-RADS 3	PI-RADS 4	PI-RADS 5	Total
No. Men	2	16 (48%)	7 (21%)	9 (27%)	1 (3%)	35
PSAD $>0,15$		3	0	2	1	6
Biopsy	1	3	2	9	1	16
No. Cancers	1	1	1	6	1	10
PSAD $>0,15$ in cancer	1	1	0	1	1	6
Cancers per PI-RADS-group	1/2 (50%)	1/16 (6%)	1/7 (14%)	6/9 (67%)	1/1 (100%)	
Significant cancers Gleason >6	1	1	1	3	1	7
Pathology GG cancer/biopsies	1/2 (50%)	1/3 (33%)	1/2 (50%)	3/9 (33%)	1/1 (100%)	

Discussion

The overarching objective of this dissertation is to assess refined MRI methods for the early detection of prostate cancer. Papers I and II contribute insights into how currently available AI-models can be employed to derive accurate PSAD values and streamline the pre-processing for fusion biopsies. Paper III evaluates and scrutinizes the use of ADC ratios in a multicenter setting, while Paper IV assesses the feasibility of a potential future screening model, including MRI, for prostate cancer.

Ad Paper I. A deep learning algorithm performs similarly to radiologists in the assessment of prostate volume on MRI

Ad Paper II. Deep learning performance on MRI prostate gland segmentation: evaluation of two commercially available algorithms compared with an expert radiologist.

Key results: The AI-model performed similarly to radiologists using ellipsoid formula in the assessment of prostate volume on MRI. Two commercially available AI-models performed accurate whole-gland prostate segmentation on a par with an expert radiologist's manual planimetry on a real-world clinical dataset.

Previous studies have evaluated prostate volume assessment by experimental deep learning algorithms against various reference standards with results consistent with our studies [82, 93, 94, 195-197]. One of the few assessments of commercially available volume calculation tools[78] raised doubts about its performance concluding that automated segmentations must be individually assessed for accuracy, as they are not always truly representative of the prostate anatomy. A strength of our study in Paper I was its distinct approach, evaluating a commercially available AI-model on diversified multicenter MRI data without pre-adaptation[196], and comparing it to the two previously established reference standards, manual planimetry, and specimen weight volume. Additionally, the size of the test set was well in line with previous studies and so-called grand challenges[78, 93, 94, 151-153, 195, 196]. Previous studies have employed various statistical methods where correlation measures[78, 94, 195] and overlay measures have been predominant[93, 196]. We have emphasized Bland Altman plots in Paper I and DICE coefficients in Paper II, as we believe that the method comparison in Paper I was better visualized by this measure by showing all datapoints. In Paper II,

our focus was on the usability of whole gland segmentations in a potential future scenario, primarily for fusion biopsies[198, 199], and therefore, we relied on various overlay measures. Despite our attempts to mimic a true clinical scenario, our datasets in Papers I and II have limitations, as there was a dominant manufacturer in our health region, and some of the participating centers were numerically dominant in the number of performed prostate MRIs. It is well established that there is variation between different readers, and we highlighted this from several perspectives. We demonstrated slightly lower interreader agreement between experienced and inexperienced radiologists performing manual planimetry[78], but the precision remained good. We showed good agreement between two experienced radiologists using the ellipsoid formula method in line with previous studies[200]. In Paper II, we demonstrated that the two tested AI-models were on par in terms of agreement shown with DICE coefficients, with several models evaluated in grand challenges and also in line with the variation seen between different radiologists[201]. For both studies, we would have preferred to compare our AI-models against public datasets, but we were hindered by the fact that the model in Paper I had already been exposed to several of the known public datasets during a training phase, which goes against the rule that training data and test data should not overlap. The same restriction applied to at least one of the AI-models in Paper II.

Implications for MRI in the early detection of prostate cancer: Our results may contribute to understanding how AI-models can be incorporated into the workflow for efficient patient selection for biopsy based on robust PSAD values, and how AI-models can serve as time-saving tools in a radiologist's preparation for targeted biopsies with MRI/Ultrasound fusion technology.

Ad Paper III. Correlation between ADC, ADC ratio, and Gleason Grade group in prostate cancer patients undergoing radical prostatectomy: Retrospective multicenter study with different MRI scanners.

Key results: In this multi-scanner cohort comprising patients with ISUP grade two and higher, no correlation between various ADC-metrics and tumor aggressivity could be identified.

The delineation between non-significant and significant cancer is pivotal for this type of study. Previous investigations have employed varied demarcations, placing ISUP 2 in either the non-significant or significant cancer group[112, 122, 128, 129, 131]. Our cohort was constrained by the stipulation that all included patients in this RALP-cohort exhibited ISUP grade 2 or higher (in line with clinical guidelines[202]). Contrary to several studies demonstrating an inverse correlation between ADC values and tumor aggressiveness [123, 133], we were unable to observe this phenomenon, with our Area Under the Curve (AUC) standing at 0.62, in contrast to values exceeding 0.9[123, 133]. The prostate MRI examinations

included in the study were conducted as part of routine clinical practice on different scanners with varying ages, field strengths, and DWI protocol parameters. This heterogeneity may have contributed to the negative results. Several comparative studies demonstrating a strong inverse correlation were single-scanner studies, eliminating certain potential sources of error[112, 133]. A previously explored approach involves normalizing the ADC value by relating it to another part of the same organ or different tissue, with some studies showing an advantage for using ADC ratios[111, 112, 122, 123], while others failed to demonstrate added value[128, 134]. In our evaluation with three different denominators for ADC ratio derivation, the AUC for ratios was lower than for absolute ADC values. The heterogeneity in our MRI dataset included not only scanner model, age, and field strength but also varying DWI protocol parameters, which, at the time of data collection, were not harmonized with prevailing PI-RADS recommendations. This included different sets of b-values for calculating ADC values, which may have partially explained our negative results[203, 204] (Le Bihan, Gudbjartsson,). We observed good inter-rater variability among radiologists performing ADC measurements, supporting the feasibility of incorporating such measurements into the clinical routine.

Implications for MRI in the early detection of prostate cancer: The utilization of diverse ADC-metrics has a role in risk stratification when evaluating focal lesions on prostate MRI. However, caution should be exercised in generalizing thresholds and criteria in a true clinical context. Customized validation based on local conditions is recommended.

Ad Paper IV. A pilot study of an organized population-based testing program for prostate cancer

Key results: The proposed model and setup for a community-based, automated, risk-adapted organized prostate cancer testing program was operationally feasible. Increasing the age for invited men may have an impact on participation rates, PSA levels, and PCa detection rates, which could affect the resources required for a larger screening program.

The participation rate in the study was slightly lower than in the national comparative prostate cancer screening studies, Göteborg 1 and 2[7, 205], but higher than that observed in STHLM3-MRI[206]. The execution of the pilot study during the Covid pandemic may have contributed to the lower participation; however, a significant observation was that the participation rate appeared to increase with the advancing age of the invited men. Furthermore, other outcomes in the diagnostic chain, such as the number of elevated PSA levels, the number of conducted prostate MRI scans, and the number of biopsies, all exhibited an upward trend with increasing age. Our levels of elevated PSA aligned with previous studies[7, 207],

and the risk-stratified algorithm based on PSA density (PSAD) and MRI outcomes resulted in biopsy avoidance for approximately half of the men with elevated PSA. In general, interpretations of outcomes regarding PI-RADS categories and cancer incidence in biopsies must be approached with caution in such a small sample size. However, trends in our data indicated slightly less negative prostate MRI results (PI-RADS <3) and a larger proportion of PI-RADS 4 compared to national screening studies[7, 206]. This prompts reflection and underscores the need for continuous calibration for radiologists and the potential for feedback from pathology[208] to avoid excessive false-positive MRI findings leading to unnecessary biopsies. Unlike the national comparative screening studies, the radiological assessments were decentralized in our pilot study, exposing each radiologist to fewer cases compared to a scenario involving centralized review in high-volume centers. Our research group is currently undertaking multiple studies with the aim of shedding light on national inter-observer variation, optimizing MRI assessments within the OPT and implementing and evaluating an individualised automated feedback loop for radiologists.

Implications for MRI in the early detection of prostate cancer: The pilot study contributes essential knowledge regarding the structuring and implementation of a potential future screening program, offering insights into how risk stratification using PSAD and MRI can help avoid unnecessary biopsies.

Conclusions

- The AI-model performed similarly to radiologists using ellipsoid formula in the assessment of prostate volume on MRI (Paper I). Two commercially available AI-models performed accurate whole-gland prostate segmentation on a par with expert radiologist manual planimetry on a real-world clinical dataset (Paper II). Our results may contribute to understanding how AI-models can be incorporated into the workflow for efficient patient selection for biopsy based on robust PSAD values and how AI-models can serve as facilitating tools in a radiologist's preparation for targeted biopsies with MRI/Ultrasound fusion technology.
- No correlation between various ADC-metrics and tumor aggressivity could be identified in our multi-scanner cohort in Paper III. Caution should be exercised in generalizing thresholds and criteria in a true clinical context. Customized validation based on local conditions is recommended.
- The pilot study in Paper IV showed good feasibility for an MRI-first based algorithm and contributes with essential knowledge regarding structuring and implementation of a potential future screening program.

Future perspectives

Several key trajectories will shape the agenda for the continuation of the various aspects of refinements in prostate MRI for early diagnosis of prostate cancer.

Over the past year, a significant breakthrough in image enhancement has been observed in clinical practice through the implementation of artificial intelligence[148] and we will see continuous improvement and development in this field.

The MRI-first concept has been firmly established, and the evaluation of AI-models for prostate cancer detection is currently underway. Recently, the initial results from the largest reader study, PI-CAI 22[166, 172], were published, wherein this retrospective comparison between AI-models and radiologists demonstrated a fractional superiority for AI. Similar studies in pathology have also been documented[209]. Going forward, there is a need for prospective evaluative studies in this domain, including other cohorts (e.g., OPT/screening with lower cancer prevalences). Equally important is the parallel expansion of purpose-built validation platforms for AI-models[210], allowing customized evaluations based on local considerations. It is intriguing to draw inspiration from breast cancer screening, validation projects, and prospective studies utilizing AI-models in the review process, as recently published[211]. The general public's and patients' acceptance of the implementation of AI-models in healthcare, particularly within the diagnostic domain, requires further investigation[212].

The significance of histological feedback for prostate MRI radiologists is well-established[208]. On a personal note, I am deeply involved at the national level in designing and implementing personalized and automated feedback loops to radiologists from biopsy outcomes[213].

The technological evolution within MRI is continuous, presenting a potential future scenario of virtual in vivo biopsies[214]. This involves extracting sufficient information from radionomics[215] and multidimensional MRI, considering tissue heterogeneity[216, 217], to obviate the need for invasive biopsies.

The optimal design of screening for prostate cancer will be the subject of intensive research in the coming decade. Among the areas requiring further exploration are understanding the most pertinent risk markers, optimizing MRI utilization, exploring the potential of AI-assisted review, and determining the intervals for re-

inviting men for screening. Insights will accrue gradually as data emerges from ongoing national screening studies, complemented by the initiation of new international initiatives[218].

Acknowledgements

Ni är många som har hjälpt mig genomföra detta doktorandprojekt och jag vill uttrycka min tacksamhet till er!

Sophia Zackrisson, min huvudhandledare. Du kombinerar skärpa, intelligens och humor och har interagerat just precis lagom mycket för att låta mig växa som forskarstudent.

Anders Bjartell, min bihandledare. En rutinerad kapten vid rodret som pekar i rätt riktning, har förmåga att destillera fram essensen och är generös med att öppna dörrar till nya spännande sammanhang.

Erik Baubeta Fridh, min bihandledare. Närvarande och arbetsam forskarkollega som guidat mig genom de första arbetena.

Kajsa Trens, sekreterare. Tack för god hjälp med både med studierelaterade frågor och andra praktikaliteter.

OPT-Skåne och den associerade forskargruppen. *Max Alterbeck, Kevin Sandeman, Johan Bengtsson, Thomas Jiborn, Despina Flondell-Sité, Jonatan Engman, Sigrid Carlsson, Anetta Bolejko, Filip Szczepankiewicz, Anna Holst, Linda Jireteg, Anna-Karin Börjedal*. Vi har bara börjat vår resa!

Johan Bengtsson, vän, medförfattare och OPT-vapendragare. Tack för din generositet och outtömliga energi, ser fram emot din disputation och fortsatt kliniskt- och forskningsbaserat samarbete!

Max Alterbeck, OPT-fundament och medförfattare. Tack för att jag fått dela doktorandresan med dig, din nyfikenhet och ditt driv kommer leda dig långt och jag hoppas vi kan fortsätta resan tillsammans!

SCROPT, Swedish Consortium of Research on OPT. *Ola Bratt* leder den månghövddade gruppen där även *Rebecka Arnsrud Godtman* och *Anna Lantz* ingår. Tack för gott samarbete över specialitetsgränserna och för generositeten att bjuda in till nationella projekt där vi tillsammans beforskar detta unika projekt.

NPCR, Nationella Prostata Cancer Registret och särskilt arbetsgruppen för diagnostikformulären. *Nina Hageman, Maria Nyberg, Gert Malmberg, Magnus Törnblom, Johan Styrke, Fredrik Jäderling, Johan Ivarsson* och *Elin Axén*. Tillsammans förbättrar vi diagnostikkedjan och skapar bättre patientsäkerhet! Tack *Ingela Franck Lissbrant* för att du generöst bjöd in till detta sammanhang!

Forskargruppen *LUCI* ledd av Sophia Zackrisson. Tack för att ni välkomnar en katt bland hermelinerna och generöst delar med er av er samlade forskarerfarenhet.

Fredrik Jäderling, vän, medförfattare och klinisk kollega. En sann och osjälvisk mentor som bjudit in mig i otaliga sammanhang. Vi delar många dimensioner och vi har bara kommit till allegrot i vår symfoni.

Jonas Wallström, barndomsvän, medförfattare och multi-kollega. Ett privilegium att få verka tillsammans med dig. Du är klok, effektiv och rolig men framförallt en ledstjärna och vän.

IT-änglarna, utan er blir det ingen forskning. *Christer Kristiansson, Björn Svensson, Henrik Granberg, Nathalie Eriksson, och Patrik Andersson*. Tack för alla samarbeten, bug-fixar och tålamod!

Sectra-kollegor, *Fredrik Lysholm, Helena Gistvik*. Nydanande samarbeten för patientens bästa!

Siemens-colleagues, *Oskar Strand, Robert Grimm, Heinrich von Busch*. Thank you for fruitful collaboration over time and a hope for continued partnerships in the future.

MIM-colleagues, *Steven Lelie and Jessica Sternisa*, I appreciate great teamwork on Paper 2 and your support in our clinical practice.

MR-sköterskorna i Helsingborg, däribland *Kristina Steinholtz, Heléne Johansson, Åsa Knutsson och Lotta Wijk-Lindgren*. Ni är klipporna som får verksamheten att fungera, alltid hjälpsamma och beredda att göra det lilla extra!

Urologkollegorna i Helsingborg, däribland *Christian Torbrand, Andreas Forsvall och Magnus Wagenius*. Samarbete på golvet när det är som bäst, gemensamt driv för patientens bästa!

Forskarskolan Helsingborg, ledd av *Niklas Nielsen och Ulla-Britt Karlsson*. En perfekt start på doktorandresan i trevligt sällskap.

Ledningen på min hemmaklinik, Verksamhetsområde Diagnostik, NVS. *Karin Fristedt, Anders Navntoft och Sven Truong*. Er generositet kan inte överskattas, ni har givit mig möjligheten att genomföra doktorandprojektet genom deltagande i forskarskolan och flexibilitet i schemaläggning för kurser och föreläsningsuppdrag. Ert ledarskap är förstklassigt!

Radiologkollegorna vid min hemmaklinik. Bra människor och bra kollegor som gör jobbet varje dag. En unik klinik att få arbeta på.

UGG-gänget på min hemmaklinik. *Torbjörn Ahl, Alexander Bodén, Carin Cronberg, Carl-Lennart Svensson, Emma Karlsson, Filip Torle, Gustav Sundström, Johanna Berg (tillfälligt utlånad), Håkan Lindvall, Jadranka Čokalić, Sofia Elmlund, Magdalena Häggström, Karin Ericsson, Petr Vanek, Fanny Nilsson, Lina*

Andersson. Min jobbfamilj och mina vänner. Utan er hade jag inte kunnat genomföra detta projekt!

Torbjörn Ahl. vän och kollega. Professionalitet och godhet personifierad, en sann förebild.

Filip Torle. Vi var två, nu är vi många. Att knäcka ett fall tillsammans med dig är min höjdpunkt på arbetsdagen. Men vi delar så många fler intressen. Stor respekt och ödmjukhet inför ditt kunnande Filip. Att inte arbeta tillsammans är för mig inget alternativ!

Gustav Sundström, vår Mr. go-to när det gäller uroradiologi och i högsta grad delaktig i denna avhandlings framsida.

Magnus Hansson, barndomsvän och forskningsmentor. Rock solid på alla plan. En förebild och vän.

Henrik Agrell, vän och kollega. Entreprenörssjäl som alltid ser en möjlighet vid motgång. En inspirerande vänskap som växer.

Urgamla vänner, lagom obenägna att förändras och stabila ankare i tillvaron, *Johan Ursberg, Emil Steuch, Mark Sonesson, Johan Sahlström* och *Jonas Ingvar.*

Stora familjen, mina föräldrar *Annika och Håkan,* som gav mig verktygen att tro på mig själv, svärföräldrarna *Lena och Olle* som är lojala och supporterande, mina systrar *Karin, Anna och Britta* som gett mig trygghet och pepp.

Lilla familjen, mina barn *Albin, Hjalmar och Klara.* Ni är det viktigaste för mig.

Åsa, min fru. Klok, varm och god. Tack.

Ekonomiskt stöd som möjliggjort min doktorandtid. Regionala ALF-medel, NVS-doktorandveckor, stipendier från Stig och Ragna Gorthons stiftelse och anslag från stiftelsen Thelma Zoégas fond för medicinsk forskning.

References

1. Cancer, I.A.f.R.o. *Cancer IAfRo. Global Cancer Observatory*. 2020 [cited 2020 2023-01-31]; Available from: <https://gco.iarc.fr/today>.
2. Welfare, S.N.B.o.H.a. *Statistics on cancer incidence*. 2021 [cited 2021 2023-01-31]; Available from: www.socialstyrelsen.se.
3. 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.
4. Kasivisvanathan, V., et al., *MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis*. *N Engl J Med*, 2018. **378**(19): p. 1767-1777.
5. Rouvière, O., et al., *Use of prostate systematic and targeted biopsy on the basis of multiparametric MRI in biopsy-naïve patients (MRI-FIRST): a prospective, multicentre, paired diagnostic study*. *Lancet Oncol*, 2019. **20**(1): p. 100-109.
6. 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.
7. Hugosson, J., et al., *Prostate Cancer Screening with PSA and MRI Followed by Targeted Biopsy Only*. *N Engl J Med*, 2022. **387**(23): p. 2126-2137.
8. EU, C., *Council recommendation of 9 December 2022 on strengthening prevention through early detection: A new EU approach on cancer screening replacing Council Recommendation 2003/878/EC.*, in *Official Journal of the European Union*. . 2022, Council EU.
9. Sweden, C.o.R.C.C.i. *Recommendations on Organised Prostate Cancer Testing (OPT)*. 2023 [cited 2023 06/09/2023]; Available from: <https://cancercentrum.se/samverkan/vara-uppdrag/prevention-och-tidig-upptackt/prostatacancertestning/organised-prostate-cancer-testing/>.
10. Aaron, L., O.E. Franco, and S.W. Hayward, *Review of Prostate Anatomy and Embryology and the Etiology of Benign Prostatic Hyperplasia*. *Urol Clin North Am*, 2016. **43**(3): p. 279-88.
11. Evatt, E.J., *A Contribution to the Development of the Prostate in Man*. *J Anat Physiol*, 1909. **43**(Pt 4): p. 314-21.
12. Lowsley, O.S., *The development of the human prostate gland with reference to the development of other structures at the neck of the urinary bladder*. *American journal of Anatomy*, 1912. **13**(3): p. 299-349.
13. Vilanova, J.C., Catalá, V., Algaba, F., & Laucirica, O., *Atlas of Multiparametric Prostate MRI: With PI-RADS Approach and Anatomic-MRI-Pathological Correlation*. 2017, Springer.

14. Engholm G, F.J., Christensen N, Kejs AMT, Johannesen TB, Khan S, et al., *NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 7.2 (16.12.2015)*. Association of the Nordic Cancer Registries: Danish Cancer Society; 2015 [updated 16.12.2015].
15. Engholm, G., et al., *NORDCAN--a Nordic tool for cancer information, planning, quality control and research*. Acta Oncol, 2010. **49**(5): p. 725-36.
16. Häggström, C., et al., *Prospective study on metabolic factors and risk of prostate cancer*. Cancer, 2012. **118**(24): p. 6199-206.
17. Mandair, D., et al., *Prostate cancer and the influence of dietary factors and supplements: a systematic review*. Nutr Metab (Lond), 2014. **11**: p. 30.
18. Moreira, D.M., et al., *Cigarette smoking is associated with an increased risk of biochemical disease recurrence, metastasis, castration-resistant prostate cancer, and mortality after radical prostatectomy: results from the SEARCH database*. Cancer, 2014. **120**(2): p. 197-204.
19. Kiciński, M., J. Vangronsveld, and T.S. Nawrot, *An epidemiological reappraisal of the familial aggregation of prostate cancer: a meta-analysis*. PLoS One, 2011. **6**(10): p. e27130.
20. Alane, S.R., et al., *Clinical features and management of BRCA1 and BRCA2-associated prostate cancer*. Front Biosci (Elite Ed), 2014. **6**(1): p. 15-30.
21. Akbari, M.R., et al., *The impact of a BRCA2 mutation on mortality from screen-detected prostate cancer*. Br J Cancer, 2014. **111**(6): p. 1238-40.
22. Stattin, P., et al., *Outcomes in localized prostate cancer: National Prostate Cancer Register of Sweden follow-up study*. J Natl Cancer Inst, 2010. **102**(13): p. 950-8.
23. Rider, J.R., et al., *Long-term outcomes among noncuratively treated men according to prostate cancer risk category in a nationwide, population-based study*. Eur Urol, 2013. **63**(1): p. 88-96.
24. Bill-Axelsson, A., et al., *Radical Prostatectomy or Watchful Waiting in Prostate Cancer - 29-Year Follow-up*. N Engl J Med, 2018. **379**(24): p. 2319-2329.
25. Bell, K.J., et al., *Prevalence of incidental prostate cancer: A systematic review of autopsy studies*. Int J Cancer, 2015. **137**(7): p. 1749-57.
26. Trpkov, K., et al., *'Insignificant' prostate cancer on prostatectomy and cystoprostatectomy: variation on a theme 'low-volume/low-grade' prostate cancer?* BJU Int, 2010. **106**(3): p. 304-15.
27. Loeb, S. and W.J. Catalona, *The Prostate Health Index: a new test for the detection of prostate cancer*. Ther Adv Urol, 2014. **6**(2): p. 74-7.
28. Woodrum, D.L., et al., *Interpretation of free prostate specific antigen clinical research studies for the detection of prostate cancer*. J Urol, 1998. **159**(1): p. 5-12.
29. Schoots, I.G. and A.R. Padhani, *Risk-adapted biopsy decision based on prostate magnetic resonance imaging and prostate-specific antigen density for enhanced biopsy avoidance in first prostate cancer diagnostic evaluation*. BJU Int, 2021. **127**(2): p. 175-178.

30. Deniffel, D., et al., *Avoiding Unnecessary Biopsy: MRI-based Risk Models versus a PI-RADS and PSA Density Strategy for Clinically Significant Prostate Cancer*. Radiology, 2021. **300**(2): p. 369-379.
31. Boesen, L., et al., *Prebiopsy Biparametric Magnetic Resonance Imaging Combined with Prostate-specific Antigen Density in Detecting and Ruling out Gleason 7-10 Prostate Cancer in Biopsy-naïve Men*. Eur Urol Oncol, 2019. **2**(3): p. 311-319.
32. EAU, *EAU guidelines*. Available via: <http://uroweb.org/guidelines/compilations-of-all-guidelines/>. 2023, EAU Guidelines Office, Arnhem, The Netherlands.
33. Centres, R.a.C.o.R.C., *Swedish National Clinical Cancer Care guidelines for prostate cancer*. Available in swedish from <https://kunksapsbanken.cancercentrum.se/diagnoser/prostatacancer/vardprogram>.
34. Thompson, I.M., et al., *Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter*. N Engl J Med, 2004. **350**(22): p. 2239-46.
35. Saini, S., *PSA and beyond: alternative prostate cancer biomarkers*. Cell Oncol (Dordr), 2016. **39**(2): p. 97-106.
36. Sumura, M., et al., *Initial evaluation of prostate cancer with real-time elastography based on step-section pathologic analysis after radical prostatectomy: a preliminary study*. Int J Urol, 2007. **14**(9): p. 811-6.
37. Gleason, D.F., *Classification of prostatic carcinomas*. Cancer Chemother Rep, 1966. **50**(3): p. 125-8.
38. Egevad, L., et al., *Handling and reporting of radical prostatectomy specimens in Europe: a web-based survey by the European Network of Uropathology (ENUP)*. Histopathology, 2008. **53**(3): p. 333-9.
39. 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.
40. 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.
41. Allsbrook, W.C., Jr., et al., *Interobserver reproducibility of Gleason grading of prostatic carcinoma: urologic pathologists*. Hum Pathol, 2001. **32**(1): p. 74-80.
42. Persson, J., et al., *Interobserver variability in the pathological assessment of radical prostatectomy specimens: findings of the Laparoscopic Prostatectomy Robot Open (LAPPRO) study*. Scand J Urol, 2014. **48**(2): p. 160-7.
43. Egevad, L., et al., *Standardization of Gleason grading among 337 European pathologists*. Histopathology, 2013. **62**(2): p. 247-56.
44. D'Amico, A.V., et al., *Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer*. Jama, 1998. **280**(11): p. 969-74.

45. Studer, U.E., et al., *Using PSA to guide timing of androgen deprivation in patients with T0-4 N0-2 M0 prostate cancer not suitable for local curative treatment (EORTC 30891)*. Eur Urol, 2008. **53**(5): p. 941-9.
46. Denham, J.W., et al., *Short-term neoadjuvant androgen deprivation and radiotherapy for locally advanced prostate cancer: 10-year data from the TROG 96.01 randomised trial*. Lancet Oncol, 2011. **12**(5): p. 451-9.
47. Widmark, A., et al., *Ultra-hypofractionated versus conventionally fractionated radiotherapy for prostate cancer: 5-year outcomes of the HYPO-RT-PC randomised, non-inferiority, phase 3 trial*. Lancet, 2019. **394**(10196): p. 385-395.
48. Steyn, J.H. and F.W. Smith, *Nuclear magnetic resonance imaging of the prostate*. Br J Urol, 1982. **54**(6): p. 726-8.
49. Schnall, M.D., et al., *The seminal tract in patients with ejaculatory dysfunction: MR imaging with an endorectal surface coil*. AJR Am J Roentgenol, 1992. **159**(2): p. 337-41.
50. Engelbrecht, M.R., et al., *Local staging of prostate cancer using magnetic resonance imaging: a meta-analysis*. Eur Radiol, 2002. **12**(9): p. 2294-302.
51. Sillerud, L.O., et al., *In vivo ¹³C NMR spectroscopy of the human prostate*. Magn Reson Med, 1988. **8**(2): p. 224-30.
52. Grant, K.B., et al., *Comparison of calculated and acquired high b value diffusion-weighted imaging in prostate cancer*. Abdom Imaging, 2015. **40**(3): p. 578-86.
53. Tamada, T., et al., *High b value (2,000 s/mm²) diffusion-weighted magnetic resonance imaging in prostate cancer at 3 Tesla: comparison with 1,000 s/mm² for tumor conspicuity and discrimination of aggressiveness*. PLoS One, 2014. **9**(5): p. e96619.
54. 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.
55. 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.
56. 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.
57. 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*. Eur Radiol, 2021. **31**(11): p. 8692-8702.
58. Rosenkrantz, A.B., et al., *Impact of delay after biopsy and post-biopsy haemorrhage on prostate cancer tumour detection using multi-parametric MRI: a multi-reader study*. Clin Radiol, 2012. **67**(12): p. e83-90.
59. Giganti, F., et al., *Prostate Imaging Quality (PI-QUAL): A New Quality Control Scoring System for Multiparametric Magnetic Resonance Imaging of the Prostate from the PRECISION trial*. Eur Urol Oncol, 2020. **3**(5): p. 615-619.
60. Caglic, I., et al., *Evaluating the effect of rectal distension on prostate multiparametric MRI image quality*. Eur J Radiol, 2017. **90**: p. 174-180.

61. Chang, S.D., et al., *Canadian Association of Radiologists Prostate MRI White Paper*. Can Assoc Radiol J, 2022. **73**(4): p. 626-638.
62. 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, 2021. **216**(1): p. 3-19.
63. Slough, R.A., et al., *Effect of hyoscine butylbromide on prostate multiparametric MRI anatomical and functional image quality*. Clin Radiol, 2018. **73**(2): p. 216.e9-216.e14.
64. Ullrich, T., et al., *Hyoscine butylbromide significantly decreases motion artefacts and allows better delineation of anatomic structures in mp-MRI of the prostate*. Eur Radiol, 2018. **28**(1): p. 17-23.
65. 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.
66. Prabhakar, S. and N. Schieda, *Patient preparation for prostate MRI: A scoping review*. Eur J Radiol, 2023. **162**: p. 110758.
67. Brennan, D.L., et al., *Do antispasmodics or rectal enemas improve image quality on multiparametric prostate MRI? An 'Evidence-Based Practice' review of the literature*. Abdom Radiol (NY), 2021. **46**(6): p. 2770-2778.
68. Lee, C., J.M. Kozlowski, and J.T. Grayhack, *Intrinsic and extrinsic factors controlling benign prostatic growth*. Prostate, 1997. **31**(2): p. 131-8.
69. Berry, S.J., et al., *The development of human benign prostatic hyperplasia with age*. J Urol, 1984. **132**(3): p. 474-9.
70. Rukstalis, D., et al., *Prostatic Urethral Lift (PUL) for obstructive median lobes: 12 month results of the MedLift Study*. Prostate Cancer Prostatic Dis, 2019. **22**(3): p. 411-419.
71. Krupski, T., et al., *The impact of prostate volume following neoadjuvant androgen deprivation on quality of life and voiding symptoms in patients undergoing permanent prostate brachytherapy*. Eur Urol, 2003. **43**(5): p. 467-72.
72. Locke, J., et al., *Risk factors for acute urinary retention requiring temporary intermittent catheterization after prostate brachytherapy: a prospective study*. Int J Radiat Oncol Biol Phys, 2002. **52**(3): p. 712-9.
73. Aizer, A.A., et al., *The impact of pretreatment prostate volume on severe acute genitourinary toxicity in prostate cancer patients treated with intensity-modulated radiation therapy*. Int J Radiat Oncol Biol Phys, 2011. **79**(2): p. 379-84.
74. Pinkawa, M., et al., *Toxicity profile with a large prostate volume after external beam radiotherapy for localized prostate cancer*. Int J Radiat Oncol Biol Phys, 2008. **70**(1): p. 83-9.
75. Benson, M.C., et al., *Prostate specific antigen density: a means of distinguishing benign prostatic hypertrophy and prostate cancer*. J Urol, 1992. **147**(3 Pt 2): p. 815-6.
76. Aminsharifi, A., et al., *Prostate Specific Antigen Density as a Predictor of Clinically Significant Prostate Cancer When the Prostate Specific Antigen is in the Diagnostic Gray Zone: Defining the Optimum Cutoff Point Stratified by Race and Body Mass Index*. J Urol, 2018. **200**(4): p. 758-766.

77. Roehrborn, C.G., et al., *Correlation between prostate size estimated by digital rectal examination and measured by transrectal ultrasound*. Urology, 1997. **49**(4): p. 548-57.
78. Bezinque, A., et al., *Determination of Prostate Volume: A Comparison of Contemporary Methods*. Acad Radiol, 2018. **25**(12): p. 1582-1587.
79. Sim, K.C., et al., *Magnetic Resonance Imaging–Based Prostate-Specific Antigen Density for Prediction of Gleason Score Upgrade in Patients With Low-Risk Prostate Cancer on Initial Biopsy*. Journal of computer assisted tomography, 2017. **41**(5): p. 731-736.
80. Karademir, I., et al., *Prostate volumes derived from MRI and volume-adjusted serum prostate-specific antigen: correlation with Gleason score of prostate cancer*. American Journal of Roentgenology, 2013. **201**(5): p. 1041-1048.
81. Wasserman, N.F., E. Niendorf, and B. Spilseth, *Measurement of Prostate Volume with MRI (A Guide for the Perplexed): Biproximate Method with Analysis of Precision and Accuracy*. Sci Rep, 2020. **10**(1): p. 575.
82. Salvaggio, G., et al., *Deep Learning Network for Segmentation of the Prostate Gland With Median Lobe Enlargement in T2-weighted MR Images: Comparison With Manual Segmentation Method*. Curr Probl Diagn Radiol, 2021.
83. Pate, W.R., et al., *Comparison of Transabdominal and Transrectal Ultrasound for Sizing of the Prostate*. Urology, 2020. **141**: p. 125-129.
84. Jones, D.R., et al., *Assessment of volume measurement of the prostate using per-rectal ultrasonography*. Br J Urol, 1989. **64**(5): p. 493-5.
85. Toth, R., et al., *Accurate prostate volume estimation using multifeature active shape models on T2-weighted MRI*. Academic radiology, 2011. **18**(6): p. 745-754.
86. Choi, Y.J., et al., *Interobserver variability of transrectal ultrasound for prostate volume measurement according to volume and observer experience*. AJR Am J Roentgenol, 2009. **192**(2): p. 444-9.
87. Ma, L., et al. *Automatic segmentation of the prostate on CT images using deep learning and multi-atlas fusion*. in *Medical Imaging 2017: Image Processing*. 2017. International Society for Optics and Photonics.
88. Alterbeck, M., et al., *Designing and Implementing a Population-based Organised Prostate Cancer Testing Programme*. Eur Urol Focus, 2022.
89. Alterbeck, M., et al., *A pilot study of an organised population-based testing programme for prostate cancer*. BJU Int, 2023.
90. Jeong, C.W., et al., *Comparison of prostate volume measured by transrectal ultrasonography and MRI with the actual prostate volume measured after radical prostatectomy*. Urologia internationalis, 2008. **81**(2): p. 179-185.
91. Thimansson, E., et al., *Deep learning algorithm performs similarly to radiologists in the assessment of prostate volume on MRI*. Eur Radiol, 2022.
92. Hong, M.K., et al., *Prostate weight is the preferred measure of prostate size in radical prostatectomy cohorts*. BJU Int, 2012. **109 Suppl 3**: p. 57-63.

93. Lee, D.K., et al., *Three-Dimensional Convolutional Neural Network for Prostate MRI Segmentation and Comparison of Prostate Volume Measurements by Use of Artificial Neural Network and Ellipsoid Formula*. AJR Am J Roentgenol, 2020. **214**(6): p. 1229-1238.
94. Turkbey, B., et al., *Fully automated prostate segmentation on MRI: comparison with manual segmentation methods and specimen volumes*. American Journal of Roentgenology, 2013. **201**(5): p. W720-W729.
95. Wan, H. *Automated Contouring Using Neural Networks Contour ProtegeAI*. 2020 [cited 2023 1/3/2023]; Available from: available via <https://go.mimsoftware.com/contour-protegeai/white-paper>.
96. Bates, T.S., et al., *Determination of prostatic volume with transrectal ultrasound: A study of intra-observer and interobserver variation*. J Urol, 1996. **155**(4): p. 1299-300.
97. Bratt, O., et al., *Population-based Organised Prostate Cancer Testing: Results from the First Invitation of 50-year-old Men*. Eur Urol, 2023.
98. Nordström, 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.
99. Dietrich, O., et al., *Technical aspects of MR diffusion imaging of the body*. Eur J Radiol, 2010. **76**(3): p. 314-22.
100. Surov, A., H.J. Meyer, and A. Wienke, *Correlation between apparent diffusion coefficient (ADC) and cellularity is different in several tumors: a meta-analysis*. Oncotarget, 2017. **8**(35): p. 59492-59499.
101. Hauge, A., et al., *Diffusion-weighted MRI-derived ADC values reflect collagen I content in PDX models of uterine cervical cancer*. Oncotarget, 2017. **8**(62): p. 105682-105691.
102. Koontz, N.A. and R.H. Wiggins, 3rd, *Differentiation of Benign and Malignant Head and Neck Lesions With Diffusion Tensor Imaging and DWI*. AJR Am J Roentgenol, 2017. **208**(5): p. 1110-1115.
103. Suo, S., et al., *Characterization of breast masses as benign or malignant at 3.0T MRI with whole-lesion histogram analysis of the apparent diffusion coefficient*. J Magn Reson Imaging, 2016. **43**(4): p. 894-902.
104. Costantini, M., et al., *Diffusion-weighted imaging in breast cancer: relationship between apparent diffusion coefficient and tumour aggressiveness*. Clin Radiol, 2010. **65**(12): p. 1005-12.
105. Wu, C.C., et al., *Predicting Genotype and Survival in Glioma Using Standard Clinical MR Imaging Apparent Diffusion Coefficient Images: A Pilot Study from The Cancer Genome Atlas*. AJNR Am J Neuroradiol, 2018. **39**(10): p. 1814-1820.
106. Sun, Y., et al., *Apparent Diffusion Coefficient (ADC) value: a potential imaging biomarker that reflects the biological features of rectal cancer*. PLoS One, 2014. **9**(10): p. e109371.
107. Zhang, J., et al., *Age-related changes of normal prostate: evaluation by MR diffusion tensor imaging*. Int J Clin Exp Med, 2015. **8**(7): p. 11220-4.

108. Jambor, I., et al., *Evaluation of different mathematical models for diffusion-weighted imaging of normal prostate and prostate cancer using high b-values: a repeatability study*. Magn Reson Med, 2015. **73**(5): p. 1988-98.
109. Pierre, T., et al., *Diffusion-weighted imaging of the prostate: should we use quantitative metrics to better characterize focal lesions originating in the peripheral zone?* Eur Radiol, 2018. **28**(5): p. 2236-2245.
110. Lucarelli, N.M., et al., *Quantitative ADC: An Additional Tool in the Evaluation of Prostate Cancer?* J Pers Med, 2023. **13**(9).
111. Barrett, T., et al., *Ratio of Tumor to Normal Prostate Tissue Apparent Diffusion Coefficient as a Method for Quantifying DWI of the Prostate*. AJR Am J Roentgenol, 2015. **205**(6): p. W585-93.
112. Boesen, L., et al., *Apparent diffusion coefficient ratio correlates significantly with prostate cancer gleason score at final pathology*. J Magn Reson Imaging, 2015. **42**(2): p. 446-53.
113. Hötker, A.M., et al., *Assessment of Prostate Cancer Aggressiveness by Use of the Combination of Quantitative DWI and Dynamic Contrast-Enhanced MRI*. AJR Am J Roentgenol, 2016. **206**(4): p. 756-63.
114. Peng, Y., et al., *Apparent diffusion coefficient for prostate cancer imaging: impact of B values*. AJR Am J Roentgenol, 2014. **202**(3): p. W247-53.
115. Bae, H., et al., *Apparent diffusion coefficient value as a biomarker reflecting morphological and biological features of prostate cancer*. Int Urol Nephrol, 2014. **46**(3): p. 555-61.
116. Bengtsson, J., et al., *Correlation between ADC, ADC ratio, and Gleason Grade group in prostate cancer patients undergoing radical prostatectomy: Retrospective multicenter study with different MRI scanners*. Front Oncol, 2023. **13**: p. 1079040.
117. Manetta, R., et al., *Correlation between ADC values and Gleason score in evaluation of prostate cancer: multicentre experience and review of the literature*. Gland Surg, 2019. **8**(Suppl 3): p. S216-s222.
118. Meyer, H.J., A. Wienke, and A. Surov, *Discrimination between clinical significant and insignificant prostate cancer with apparent diffusion coefficient - a systematic review and meta analysis*. BMC Cancer, 2020. **20**(1): p. 482.
119. Merisaari, H., et al., *Diffusion-weighted imaging of prostate cancer: effect of b-value distribution on repeatability and cancer characterization*. Magn Reson Imaging, 2015. **33**(10): p. 1212-1218.
120. Thörmer, G., et al., *Diagnostic value of ADC in patients with prostate cancer: influence of the choice of b values*. Eur Radiol, 2012. **22**(8): p. 1820-8.
121. Møller, J.M., et al., *Quantification of cross-vendor variation in ADC measurements in vendor-specific prostate MRI-protocols*. Eur J Radiol, 2023. **165**: p. 110942.
122. Lebovici, A., et al., *Evaluation of the normal-to-diseased apparent diffusion coefficient ratio as an indicator of prostate cancer aggressiveness*. BMC Med Imaging, 2014. **14**: p. 15.

123. Litjens, G.J., et al., *Interpatient variation in normal peripheral zone apparent diffusion coefficient: effect on the prediction of prostate cancer aggressiveness*. Radiology, 2012. **265**(1): p. 260-6.
124. Poretti, A., et al., *Apparent diffusion coefficient of pediatric cerebellar tumors: a biomarker of tumor grade?* Pediatr Blood Cancer, 2013. **60**(12): p. 2036-41.
125. Chawalparit, O., et al., *Diagnostic performance of advanced MRI in differentiating high-grade from low-grade gliomas in a setting of routine service*. J Med Assoc Thai, 2013. **96**(10): p. 1365-73.
126. Park, S.O., et al., *Relative apparent diffusion coefficient: determination of reference site and validation of benefit for detecting metastatic lymph nodes in uterine cervical cancer*. J Magn Reson Imaging, 2009. **29**(2): p. 383-90.
127. Alessandrino, F., et al., *Predictive role of PI-RADSv2 and ADC parameters in differentiating Gleason pattern 3 + 4 and 4 + 3 prostate cancer*. Abdom Radiol (NY), 2019. **44**(1): p. 279-285.
128. Woo, S., et al., *Preoperative Evaluation of Prostate Cancer Aggressiveness: Using ADC and ADC Ratio in Determining Gleason Score*. AJR Am J Roentgenol, 2016. **207**(1): p. 114-20.
129. Bajgirani, A.M., et al., *Apparent Diffusion Coefficient (ADC) Ratio Versus Conventional ADC for Detecting Clinically Significant Prostate Cancer With 3-T MRI*. AJR Am J Roentgenol, 2019. **213**(3): p. W134-w142.
130. Falaschi, Z., et al., *Accuracy of ADC ratio in discriminating true and false positives in multiparametric prostatic MRI*. Eur J Radiol, 2020. **128**: p. 109024.
131. Pepe, P., et al., *Multiparametric MRI Apparent Diffusion Coefficient (ADC) Accuracy in Diagnosing Clinically Significant Prostate Cancer*. In Vivo, 2017. **31**(3): p. 415-418.
132. Vargas, H.A., et al., *Diffusion-weighted endorectal MR imaging at 3 T for prostate cancer: tumor detection and assessment of aggressiveness*. Radiology, 2011. **259**(3): p. 775-84.
133. De Cobelli, F., et al., *Apparent diffusion coefficient value and ratio as noninvasive potential biomarkers to predict prostate cancer grading: comparison with prostate biopsy and radical prostatectomy specimen*. AJR Am J Roentgenol, 2015. **204**(3): p. 550-7.
134. Rosenkrantz, A.B., et al., *Prostate cancer vs. post-biopsy hemorrhage: diagnosis with T2- and diffusion-weighted imaging*. J Magn Reson Imaging, 2010. **31**(6): p. 1387-94.
135. Winkel, D.J.e.a., *A Fully Automated, End-to-End Prostate MRI Workflow Solution Incorporating Dot, Ultrashort Biparametric Imaging and DeepLearning-based Detection, Classification, and Reporting*. 2020, Siemens Healthineers: Magnetom Flash (76) 1/202, available via: <https://www.magnetomworld.siemens-healthineers.com/publications/magnetom-flash>
136. Allen, B., et al., *Evaluation and Real-World Performance Monitoring of Artificial Intelligence Models in Clinical Practice Purchase: Try It, Buy It, Check It*. J Am Coll Radiol, 2021.
137. Angus, D.C., *Randomized Clinical Trials of Artificial Intelligence*. Jama, 2020. **323**(11): p. 1043-1045.

138. van Leeuwen, K.G., et al., *Artificial intelligence in radiology: 100 commercially available products and their scientific evidence*. Eur Radiol, 2021. **31**(6): p. 3797-3804.
139. Bardis, M.D., et al., *Applications of Artificial Intelligence to Prostate Multiparametric MRI (mpMRI): Current and Emerging Trends*. Cancers (Basel), 2020. **12**(5).
140. Li, H., et al., *Machine Learning in Prostate MRI for Prostate Cancer: Current Status and Future Opportunities*. Diagnostics (Basel), 2022. **12**(2).
141. Suarez-Ibarrola, R., et al., *Artificial Intelligence in Magnetic Resonance Imaging-based Prostate Cancer Diagnosis: Where Do We Stand in 2021?* Eur Urol Focus, 2021.
142. Sunoqrot, M.R.S., et al., *Artificial intelligence for prostate MRI: open datasets, available applications, and grand challenges*. Eur Radiol Exp, 2022. **6**(1): p. 35.
143. Turkbey, B. and M.A. Haider, *Deep learning-based artificial intelligence applications in prostate MRI: brief summary*. Br J Radiol, 2022. **95**(1131): p. 20210563.
144. Twilt, J.J., et al., *Artificial Intelligence Based Algorithms for Prostate Cancer Classification and Detection on Magnetic Resonance Imaging: A Narrative Review*. Diagnostics (Basel), 2021. **11**(6).
145. Penzkofer, T., et al., *ESUR/ESUI position paper: developing artificial intelligence for precision diagnosis of prostate cancer using magnetic resonance imaging*. Eur Radiol, 2021. **31**(12): p. 9567-9578.
146. Kim, B., et al., *Deep learning-based imaging reconstruction for MRI after neoadjuvant chemoradiotherapy for rectal cancer: effects on image quality and assessment of treatment response*. Abdom Radiol (NY), 2023. **48**(1): p. 201-210.
147. Park, J.C., et al., *Fast T2-Weighted Imaging With Deep Learning-Based Reconstruction: Evaluation of Image Quality and Diagnostic Performance in Patients Undergoing Radical Prostatectomy*. J Magn Reson Imaging, 2022. **55**(6): p. 1735-1744.
148. Wang, X., et al., *Novel deep learning-based noise reduction technique for prostate magnetic resonance imaging*. Abdom Radiol (NY), 2021. **46**(7): p. 3378-3386.
149. Knoll, F., et al., *Deep-Learning Methods for Parallel Magnetic Resonance Imaging Reconstruction: A Survey of the Current Approaches, Trends, and Issues*. IEEE Signal Process Mag, 2020. **37**(1): p. 128-140.
150. Lin, D.J., et al., *Artificial Intelligence for MR Image Reconstruction: An Overview for Clinicians*. J Magn Reson Imaging, 2021. **53**(4): p. 1015-1028.
151. Litjens, G., et al., *Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge*. Med Image Anal, 2014. **18**(2): p. 359-73.
152. Challenge, N.-I. *NCI-ISBI 2013 Challenge - Automated Segmentation of Prostate Structures (2015)*. 2015 [cited 2023 01/03/2023]; Available from: <https://wiki.cancerimagingarchive.net/display/Public/NCI-ISBI+2013+Challenge+-+Automated+Segmentation+of+Prostate+Structures>.
153. Antonelli, M., et al., *The Medical Segmentation Decathlon*. Nat Commun, 2022. **13**(1): p. 4128.
154. Milletari, F., N. Navab, and S.-A. Ahmadi. *V-net: Fully convolutional neural networks for volumetric medical image segmentation*. in 2016 fourth international conference on 3D vision (3DV). 2016. IEEE.

155. Cuocolo, R., et al., *Machine learning applications in prostate cancer magnetic resonance imaging*. European radiology experimental, 2019. **3**(1): p. 1-8.
156. Mun, J., et al. *Comparison of objective functions in CNN-based prostate magnetic resonance image segmentation*. in *2017 IEEE International Conference on Image Processing (ICIP)*. 2017. IEEE.
157. Tian, Z., L. Liu, and B. Fei, *Deep convolutional neural network for prostate MR segmentation*. International journal of computer assisted radiology and surgery, 2018. **13**(11): p. 1687.
158. Cheng, R., et al., *Fully automated prostate whole gland and central gland segmentation on MRI using holistically nested networks with short connections*. J Med Imaging (Bellingham), 2019. **6**(2): p. 024007.
159. Sanford, T.H., et al., *Data Augmentation and Transfer Learning to Improve Generalizability of an Automated Prostate Segmentation Model*. AJR Am J Roentgenol, 2020. **215**(6): p. 1403-1410.
160. Turkbey, B., et al., *Fully automated prostate segmentation on MRI: comparison with manual segmentation methods and specimen volumes*. AJR Am J Roentgenol, 2013. **201**(5): p. W720-9.
161. Wang, B., et al., *Deeply supervised 3D fully convolutional networks with group dilated convolution for automatic MRI prostate segmentation*. Med Phys, 2019. **46**(4): p. 1707-1718.
162. Zabihollahy, F., et al., *Automated segmentation of prostate zonal anatomy on T2-weighted (T2W) and apparent diffusion coefficient (ADC) map MR images using U-Nets*. Med Phys, 2019. **46**(7): p. 3078-3090.
163. Arif, M., et al., *Clinically significant prostate cancer detection and segmentation in low-risk patients using a convolutional neural network on multi-parametric MRI*. Eur Radiol, 2020. **30**(12): p. 6582-6592.
164. Ishioka, J., et al., *Computer-aided diagnosis of prostate cancer on magnetic resonance imaging using a convolutional neural network algorithm*. BJU Int, 2018. **122**(3): p. 411-417.
165. Roest, C., et al., *AI-assisted biparametric MRI surveillance of prostate cancer: feasibility study*. Eur Radiol, 2023. **33**(1): p. 89-96.
166. Saha A, B.J., Twilt J, et al, *Open review: Artificial Intelligence and Radiologists at Prostate Cancer Detection in MRI – The PI-CAI Challenge*. . 2023, MIDL 2023 Submission.
167. Schelb, P., et al., *Classification of Cancer at Prostate MRI: Deep Learning versus Clinical PI-RADS Assessment*. Radiology, 2019. **293**(3): p. 607-617.
168. Yoo, S., et al., *Prostate Cancer Detection using Deep Convolutional Neural Networks*. Sci Rep, 2019. **9**(1): p. 19518.
169. Shah, U., et al., *Recent Developments in Artificial Intelligence-Based Techniques for Prostate Cancer Detection: A Scoping Review*. Stud Health Technol Inform, 2022. **289**: p. 268-271.

170. Winkel, D.J., et al., *A Novel Deep Learning Based Computer-Aided Diagnosis System Improves the Accuracy and Efficiency of Radiologists in Reading Biparametric Magnetic Resonance Images of the Prostate: Results of a Multireader, Multicase Study*. Invest Radiol, 2021. **56**(10): p. 605-613.
171. Sarma, K.V., et al., *Federated learning improves site performance in multicenter deep learning without data sharing*. J Am Med Inform Assoc, 2021. **28**(6): p. 1259-1264.
172. Saha, A., Twilt, J. J., Bosma, J et al, *Artificial Intelligence and Radiologists at Prostate Cancer Detection in MRI: The PI-CAI Challenge (Study Protocol) (1.1)*. Zenodo. Zenodo.
173. Wilson, J.M.G., Jungner, Gunnar & World Health Organization, *Principles and practice of screening for disease*. Available via: <https://iris.who.int/handle/10665/37650>. 1968.
174. Ebell, M.H., T.N. Thai, and K.J. Royalty, *Cancer screening recommendations: an international comparison of high income countries*. Public Health Rev, 2018. **39**: p. 7.
175. Heijnsdijk, E.A., et al., *Overdetection, overtreatment and costs in prostate-specific antigen screening for prostate cancer*. Br J Cancer, 2009. **101**(11): p. 1833-8.
176. Loeb, S., et al., *Overdiagnosis and overtreatment of prostate cancer*. Eur Urol, 2014. **65**(6): p. 1046-55.
177. Stattin, P., et al., *Prostate cancer mortality in areas with high and low prostate cancer incidence*. J Natl Cancer Inst, 2014. **106**(3): p. dju007.
178. Hugosson, J., et al., *Eighteen-year follow-up of the Göteborg Randomized Population-based Prostate Cancer Screening Trial: effect of sociodemographic variables on participation, prostate cancer incidence and mortality*. Scand J Urol, 2018. **52**(1): p. 27-37.
179. Kilpeläinen, T.P., et al., *Prostate Cancer and Socioeconomic Status in the Finnish Randomized Study of Screening for Prostate Cancer*. Am J Epidemiol, 2016. **184**(10): p. 720-731.
180. Bratt, O., et al., *Screening for prostate cancer: evidence, ongoing trials, policies and knowledge gaps*. BMJ Oncology, 2023. **2**(1): p. e000039.
181. Moore, C.M., et al., *Prevalence of MRI lesions in men responding to a GP-led invitation for a prostate health check: a prospective cohort study*. BMJ Oncology, 2023. **2**(1): p. e000057.
182. SBU, S.A.f.H.T.A.a.A.o.S.S., *SBU publikationer, available via: <https://www.sbu.se/sv/publikationssok/>*.
183. SAPEA, S.A.f.P.b.E.A., *Cancer screening in Europe. Expert workshop 1*. 2022: <https://sapea.info/wp-content/uploads/cancer-screening-workshop-report-01.pdf>.
184. Bland, J.M. and D.G. Altman, *Statistical methods for assessing agreement between two methods of clinical measurement*. Lancet, 1986. **1**(8476): p. 307-10.
185. Dice, L.R., *Measures of the Amount of Ecologic Association Between Species*. Ecology, 1945. **26**(3): p. 297-302.
186. Spearman, C., *The Proof and Measurement of Association between Two Things*. The American Journal of Psychology.

187. Martin, R.M., et al., *Effect of a Low-Intensity PSA-Based Screening Intervention on Prostate Cancer Mortality: The CAP Randomized Clinical Trial*. Jama, 2018. **319**(9): p. 883-895.
188. Wickström A, Z.K., Morberg Jämterud S, *Screeningens mångsidighet – dess möjligheter och utmaningar*. 2022: Nordic Academic Press.
189. Bretthauer, M., et al., *Estimated Lifetime Gained With Cancer Screening Tests: A Meta-Analysis of Randomized Clinical Trials*. JAMA Internal Medicine, 2023. **183**(11): p. 1196-1203.
190. O, H., *Tänkande maskiner: Den artificiella intelligensens genombrott*. 2021: Fri Tanke Förlag.
191. Kurzweil, R., *The Singularity Is Near - When humans transcend biology*. 2006: Duckworth.
192. Cotra, A., *2020 Draft Report on Biological Ancors*. Available via: <https://drive.google.com/drive/u/0/folders/15ArhEPZSTYU8f012bs6ehPS6-xmhtBPP>. 2020.
193. al, Z.B.e., *Forecasting AI Progress: Evidence From a Survey Of Machine Learning Researchers*. 2022, Cornell University.
194. N, B., *Ethical Issues in Advanced Artificial Intelligence*. Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence, 2003. **2**.
195. Bulman, J.C., et al., *Automated computer-derived prostate volumes from MR imaging data: comparison with radiologist-derived MR imaging and pathologic specimen volumes*. Radiology, 2012. **262**(1): p. 144-51.
196. Cuocolo, R., et al., *Deep Learning Whole-Gland and Zonal Prostate Segmentation on a Public MRI Dataset*. J Magn Reson Imaging, 2021. **54**(2): p. 452-459.
197. Mazaheri, Y., et al., *Comparison of prostate volume measured by endorectal coil MRI to prostate specimen volume and mass after radical prostatectomy*. Acad Radiol, 2015. **22**(5): p. 556-62.
198. Soerensen, S.J.C., et al., *Deep Learning Improves Speed and Accuracy of Prostate Gland Segmentations on Magnetic Resonance Imaging for Targeted Biopsy*. J Urol, 2021. **206**(3): p. 604-612.
199. van Sloun, R.J.G., et al., *Deep Learning for Real-time, Automatic, and Scanner-adapted Prostate (Zone) Segmentation of Transrectal Ultrasound, for Example, Magnetic Resonance Imaging-transrectal Ultrasound Fusion Prostate Biopsy*. Eur Urol Focus, 2021. **7**(1): p. 78-85.
200. Ghafoor, S., et al., *Comparison of PI-RADS Versions 2.0 and 2.1 for MRI-based Calculation of the Prostate Volume*. Acad Radiol, 2020.
201. Becker, A.S., et al., *Variability of manual segmentation of the prostate in axial T2-weighted MRI: A multi-reader study*. Eur J Radiol, 2019. **121**: p. 108716.
202. Cornford, P., et al., *EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer. Part II-2020 Update: Treatment of Relapsing and Metastatic Prostate Cancer*. Eur Urol, 2021. **79**(2): p. 263-282.
203. Gudbjartsson, H. and S. Patz, *The Rician distribution of noisy MRI data*. Magn Reson Med, 1995. **34**(6): p. 910-4.

204. Le Bihan, D., *What can we see with IVIM MRI?* Neuroimage, 2019. **187**: p. 56-67.
205. Hugosson, J., et al., *Mortality results from the Göteborg randomised population-based prostate-cancer screening trial.* Lancet Oncol, 2010. **11**(8): p. 725-32.
206. Nordström, T., et al., *Prostate cancer screening using a combination of risk-prediction, MRI, and targeted prostate biopsies (STHLM3-MRI): a prospective, population-based, randomised, open-label, non-inferiority trial.* Lancet Oncol, 2021. **22**(9): p. 1240-1249.
207. Rannikko, A., et al., *Population-based randomized trial of screening for clinically significant prostate cancer ProScreen: a pilot study.* BJU Int, 2022. **130**(2): p. 193-199.
208. de Rooij, M., et al., *ESUR/ESUI consensus statements on multi-parametric MRI for the detection of clinically significant prostate cancer: quality requirements for image acquisition, interpretation and radiologists' training.* Eur Radiol, 2020. **30**(10): p. 5404-5416.
209. Bulten, W., et al., *Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge.* Nat Med, 2022. **28**(1): p. 154-163.
210. Cossío, F., et al., *VAI-B: a multicenter platform for the external validation of artificial intelligence algorithms in breast imaging.* J Med Imaging (Bellingham), 2023. **10**(6): p. 061404.
211. Lång, K., et al., *Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study.* Lancet Oncol, 2023. **24**(8): p. 936-944.
212. Hemphill, S., et al., *The implementation of artificial intelligence in radiology: a narrative review of patient perspectives.* Future Healthc J, 2023. **10**(1): p. 63-68.
213. Sweden, T.N.P.C.R.o., *Helsingborgs lasarett testar en ny digital lösning vid prostatacancerutredning.* available via <https://npcr.se/>. 2023.
214. Nicoletti, G., et al., *Virtual biopsy in prostate cancer: can machine learning distinguish low and high aggressive tumors on MRI?* Annu Int Conf IEEE Eng Med Biol Soc, 2021. **2021**: p. 3374-3377.
215. Chiacchio, G., et al., *Radiomics vs radiologist in prostate cancer. Results from a systematic review.* World J Urol, 2023. **41**(3): p. 709-724.
216. Jelescu, I.O., et al., *Challenges for biophysical modeling of microstructure.* J Neurosci Methods, 2020. **344**: p. 108861.
217. Langbein, B.J., et al., *A Pilot Study of Multidimensional Diffusion MRI for Assessment of Tissue Heterogeneity in Prostate Cancer.* Invest Radiol, 2021. **56**(12): p. 845-853.
218. Van Poppel, H., M.J. Roobol, and A. Chandran, *Early Detection of Prostate Cancer in the European Union: Combining Forces with PRAISE-U.* Eur Urol, 2023. **84**(6): p. 519-522.

Paper I





Deep learning algorithm performs similarly to radiologists in the assessment of prostate volume on MRI

Erik Thimansson^{1,2} · J. Bengtsson^{3,4,5} · E. Baubeta^{1,4,5} · J. Engman^{1,4,5} · D. Flondell-Sité^{6,7} · A. Bjartell^{6,7} · S. Zackrisson^{1,4,5}

Received: 12 April 2022 / Revised: 26 September 2022 / Accepted: 13 October 2022
© The Author(s) 2022, corrected publication 2022

Abstract

Objectives Prostate volume (PV) in combination with prostate specific antigen (PSA) yields PSA density which is an increasingly important biomarker. Calculating PV from MRI is a time-consuming, radiologist-dependent task. The aim of this study was to assess whether a deep learning algorithm can replace PI-RADS 2.1 based ellipsoid formula (EF) for calculating PV.

Methods Eight different measures of PV were retrospectively collected for each of 124 patients who underwent radical prostatectomy and preoperative MRI of the prostate (multicenter and multi-scanner MRI's 1.5 and 3 T). Agreement between volumes obtained from the deep learning algorithm (PV_{DL}) and ellipsoid formula by two radiologists (PV_{EF1} and PV_{EF2}) was evaluated against the reference standard PV obtained by manual planimetry by an expert radiologist (PV_{MPE}). A sensitivity analysis was performed using a prostatectomy specimen as the reference standard. Inter-reader agreement was evaluated between the radiologists using the ellipsoid formula and between the expert and inexperienced radiologists performing manual planimetry.

Results PV_{DL} showed better agreement and precision than PV_{EF1} and PV_{EF2} using the reference standard PV_{MPE} (mean difference [95% limits of agreement] PV_{DL}: -0.33 [-10.80; 10.14], PV_{EF1}: -3.83 [-19.55; 11.89], PV_{EF2}: -3.05 [-18.55; 12.45]) or the PV determined based on specimen weight (PV_{DL}: -4.22 [-22.52; 14.07], PV_{EF1}: -7.89 [-30.50; 14.73], PV_{EF2}: -6.97 [-30.13; 16.18]). Inter-reader agreement was excellent between the two experienced radiologists using the ellipsoid formula and was good between expert and inexperienced radiologists performing manual planimetry.

Conclusion Deep learning algorithm performs similarly to radiologists in the assessment of prostate volume on MRI.

Key Points

- A commercially available deep learning algorithm performs similarly to radiologists in the assessment of prostate volume on MRI.
- The deep-learning algorithm was previously untrained on this heterogeneous multicenter day-to-day practice MRI data set.

Keywords Magnetic resonance imaging · Prostate neoplasms · Deep learning · Prostate-specific antigen

Abbreviations

AI	Artificial intelligence	PSA	Prostate-specific antigen
ICC	Intraclass correlation	PSAD	Prostate-specific antigen density
MRI	Magnetic resonance imaging	PV	Prostate volume
PIRADS	Prostate Imaging Reporting and Data System	PVDL	Prostate volume obtained from deep learning algorithm

✉ Erik Thimansson
per_erik.thimansson@med.lu.se

¹ Department of Translational Medicine, Diagnostic Radiology, Lund University, Carl-Bertil Laurells gata 9, SE-205 02 Malmö, Sweden

² Department of Radiology, Helsingborg Hospital, Helsingborg, Sweden

³ Department of Clinical Sciences, Diagnostic Radiology, Lund University, Lund, Sweden

⁴ Department of Imaging and Functional Medicine, Skåne University Hospital, Malmö, Sweden

⁵ Department of Imaging and Functional Medicine, Skåne University Hospital, Lund, Sweden

⁶ Department of Translational Medicine, Urological Cancers, Lund University, Malmö, Sweden

⁷ Department of Urology, Skåne University Hospital, Malmö, Sweden

PVEF	Prostate volume obtained from radiologist using ellipsoid formula
PVMPE	Prostate volume obtained from manual planimetry by expert radiologist
PVMPU	Prostate volume obtained from manual planimetry by unexperienced radiologist
PVSD	Prostate volume obtained from specimen dimensions
PVSW	Prostate volume obtained from specimen weight
PVTRUS	Prostate volume obtained from transrectal ultrasound
EF	PI-RADS 2.1 based ellipsoid formula

Introduction

Prostate volume (PV) is an important parameter in the workup of benign and malignant prostate diseases [1–3]. Combining the prostate specific antigen (PSA) value and PV yields the PSA density (PSAD) [4–6]. A higher PSAD, often using a threshold of 0.15 ng/ml², indicates a higher risk of prostate cancer [7, 8]. PSAD is an increasingly important factor in making decisions on which patients will undergo biopsies [4]. Last years' paradigm shift towards "MRI first" [9, 10], meaning that the patient undergoes magnetic resonance imaging (MRI) before biopsies, has made MRI a cornerstone for determining PV. PIRADS [11] recommends the ellipsoid formula method for determined PV; the radiologist measures the prostate height, depth, and width. This ellipsoid formula (EF) is considered the gold standard and has been shown to be relatively accurate [12], but it has some limitations. Use of the EF is time-consuming, reader-dependent, and prone to multiplication errors due to the prostate not being a symmetrical geometrical ellipsoid body, posing anatomical challenges delineating the apex and ventral portion [4, 13, 14]. The most accurate method for assessing PV on MRI is manual planimetry [12, 15, 16], in which the radiologist uses external software to manually outline the prostate boundaries on T2-weighted MRI in three planes. However, manual planimetry is too time-consuming to be a realistic alternative in clinical routine [17].

In the last few years, there has been growing interest in the development of artificial intelligence (AI)-based algorithms in radiology. The most commonly used AI method in imaging is deep learning convolutional neural network-based algorithms [18]. Several studies have shown good performance of AI for automated assisted PV assessment [18–20], but questions remain about external validation, generalizability, and how well the algorithm performs in a different clinical context with heterogeneous data. The number of algorithms cleared by the U.S. Food and Drug Administration (FDA) continues to grow [21], though with increasing concern regarding algorithms' true performance in the clinical setting beyond the institutions in which they were trained and validated. To subsidize to these issues,

we designed this multicenter, multi-scanner study. We used a proprietary, commercially available deep learning-based system [22] not previously exposed to the data set. To make the comparison between PV methods as comprehensive as possible, we compared them to PVs from transrectal ultrasound and prostatectomy specimens [23–25].

The primary aim of this study was to assess whether a previously unexposed deep learning algorithm can replace the PI-RADS 2.1-based ellipsoid formula for calculating the PV from a heterogeneous data set from prostate MRI.

The secondary aims were to evaluate the inter-reader agreement between two radiologists using PI-RADS 2.1-based ellipsoid formula and between experienced and inexperienced radiologists performing manual planimetry.

Materials and methods

Study design and population

This retrospective multicenter study was approved by the local ethics review committee at Lund University (entry no. 2014-886) and the Swedish Ethical Review Authority (entry no. 2019-03674). All consecutive patients who underwent robot-assisted radical prostatectomy at Malmö University Hospital in 2018 were identified and assessed for eligibility. Patients were included if they had undergone MRI of the prostate less than 1 year before the surgery. Two patients were excluded due to MRI at a hospital outside our health care region and patient withdrawal, resulting in the inclusion of 124 patients in the study. The data collection algorithm is presented in Fig. 1. Eight different PVs were calculated per patient (Table 1).

MRI technique

The MRI scans were performed at seven different hospitals using eight different scanners, comprising seven different scanner models from two vendors, two different field strengths (1.5 and 3T), and two different T2 transaxial (axial) slice angulations. Different imaging acquisition parameters were used at different sites according to local routines. All protocols included transversal, coronal, and sagittal T2-weighted turbo spin-echo images, which were used for ellipsoid formula calculations, and the T2 axial, which was used for manual planimetry and deep learning contouring. The parameters for the T2 axial are listed in Supplemental Table 1 in electronic supplementary material.

Prostate volume calculations from imaging

All MRI exams were retrospectively read by three consulting radiologists (E.T., J.B., and J.E.) with at least 5 years of experience with MRI prostate. One radiologist (J.E.; highly

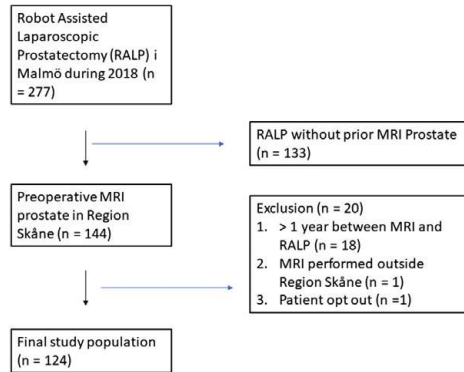


Fig. 1 Study cohort

experienced in manual planimetry from 5 years of fusion biopsy planning at a tertiary referral center) performed manual planimetry by manually tracing the prostate boundaries on the T2 image in three planes using external software (MIM Software, Inc.), blinded to all other volume calculations. This volume, PV_{MPE} , was used as the reference standard, and time consumption for the whole workflow for this manual planimetry was measured on part of the exams (14 patients).

Another radiologist (E.T.) inexperienced in manual planimetry performed manual planimetry (PV_{MPU}) using the same software, but on a different server to secure blinding. Two radiologists (E.T. and J.B.) calculated the PVs using the ellipsoid formula method ($[\text{width} \times \text{depth} \times \text{height}] \times [\pi/6]$) according to PIRADS [11], shown in Fig. 2b, c. This was done independently and blinded for all other PVs; these volumes are abbreviated PV_{EF1} and PV_{EF2} .

PVs were also calculated from transrectal ultrasound according to the ellipsoid formula. The ultrasounds were obtained during a routine clinical visit to the urology department and collected from the patient records.

Deep learning algorithm

The algorithm used is a proprietary commercially available product (AI-Rad Companion Prostate MR VA20A_HF02, Siemens Healthcare AG), a machine learning deep learning algorithm that uses a convolutional neural network deep image to image (Dl2IN) network. The algorithm was not previously exposed to the images in the current study cohort and whole gland segmentation was performed on non-annotated T2 axial images as described by Yang [26]. The contours and volume calculations (PV_{DL}) were exported back to and saved in the Picture Archive and Communications System (Sectra IDS7). The results were presented as burnt-in contours and PV as text. Contours outlined by deep learning algorithm and expert radiologist are shown in Fig. 2a.

Prostatectomy specimens

Prostatectomy specimens were processed and prepared according to international standard pathological procedures [27], embedding all material, including seminal vesicles and extraprostatic tissue. Specimen dimensions in three planes and weight were obtained from pathology reports. Specimen volume was calculated using the ellipsoid formula (PV_{SD}) [12]. Specimen weight and the prostate density coefficient (1.05 g/mL) [24, 25] were also used to calculate PV (PV_{SW}).

Statistical analysis

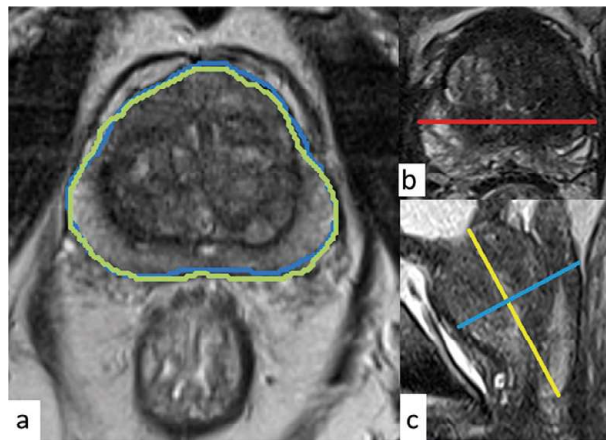
Descriptive statistics were used to describe the study cohort as medians and ranges. A box plot was used to present the distribution of volumes according to the different methods.

Agreement between volume measurement methods was evaluated using Bland Altman plots. First, we compared deep learning and ellipsoid formula-determined volumes (i.e., PV_{DL} vs. PV_{EF1} and PV_{EF2}) in relation to the manual planimetry standard (i.e., PV_{MPE}). Second, we performed a sensitivity analysis comparing the deep learning and ellipsoid

Table 1 Descriptions and abbreviations of prostate volumes

Abbreviation	Name	Description
PV_{MPE}	Prostate volume by manual planimetry, expert	Radiologist experienced in manual planimetry
PV_{MPU}	Prostate volume by manual planimetry, inexperienced	Radiologist inexperienced in manual planimetry
PV_{EF1}	Prostate volume by ellipsoid formula radiologist 1	Radiologist
PV_{EF2}	Prostate volume by ellipsoid formula radiologist 2	Radiologist
PV_{TRUS}	Prostate volume by transrectal ultrasound	Urologist
PV_{DL}	Prostate volume by deep learning	Deep learning algorithm
PV_{SD}	Prostate volume based on specimen dimensions	Measurements by pathology department staff
PV_{SW}	Prostate volume based on specimen weight	Measurements by pathology department staff

Fig. 2 **a** One out of approximately 30 T2 transaxial images with the prostate contour outlined by planimetry (blue line, expert radiologist [PV_{MPE}]; green line, deep learning algorithm [PV_{DL}]). **b** T2 transaxial and **c** T2 sagittal images with measurements for the ellipsoid formula (PV_{EF1}, PV_{EF2}; red line, transverse diameter; yellow line, longitudinal diameter; blue line, antero-posterior diameter)



formula-determined volumes in relation to the volumes determined based on specimen weight (i.e., PV_{SW}).

Using Bland Altman plots, inter-reader agreement was analyzed between an experienced and inexperienced radiologist performing manual planimetry (i.e., PV_{MPE} vs. PV_{MPI}) and between two radiologists using the ellipsoid formula for volume assessment (i.e., PV_{EF1} vs. PV_{EF2}). Inter-reader agreement was also measured for the latter using the intraclass correlation (ICC). The formula for random effects, absolute agreement, and single rater measurements was used. The paired *t*-test was used to compare the mean differences between the experienced and inexperienced radiologist performing manual planimetry and paired *t*-test with Bonferroni correction was used to compare volume methods. All statistical analyses were performed in R version 4.0.2 [28].

Results

The cohort consisted of 124 patients with a median age of 66 years (range 45–76 years) and median preoperative PSA of 6.90 µg/L (min 0.88; max 39).

As shown in Fig. 3a, the observed mean difference between PV_{DL} and PV_{MPE} was lower than the observed mean difference between PV_{EF} and PV_{MPE} (mean difference [95% limits of agreement] PV_{DL}: −0.33 mL [−10.80; 10.14 mL], PV_{EF1}: −3.83 mL [−19.55; 11.89 mL], PV_{EF2}: −3.05 mL [−18.55; 12.45 mL]). The limits of agreement were slightly narrower for PV_{DL} than PV_{EF}, indicating better precision.

A sensitivity analysis using PV_{SW} as the reference standard is presented as a Bland Altman plot in Fig. 3b. The mean

difference (bias) was lower for PV_{DL} than PV_{EF1} or PV_{EF2} and the corresponding 95% limits of agreement (mean + 1.96 SD and mean − 1.96 SD) narrower for PV_{DL} than PV_{EF1} or PV_{EF2} using PV_{MPE} as the reference standard (mean difference [95% limits of agreement] PV_{DL}: −4.22 mL [−22.52; 14.07 mL], PV_{EF1}: −7.89 mL [−30.50; 14.73 mL], PV_{EF2}: −6.97 mL [−30.13; 16.18 mL]). In both Bland Altman plots (Fig. 3a, b), there was a tendency of deep learning and the ellipsoid formula to underestimate volumes of enlarges prostates compared to both manual planimetry and specimen weight, but deep learning seemed to underestimate the large volumes to a lesser extent than the ellipsoid formula.

The inter-reader agreement between the two radiologists performing manual planimetry is shown in Fig. 4a, indicating that the inexperienced reader systematically calculated lower volumes than the experienced reader, but with better precision than between the two radiologists using the ellipsoid formula (mean difference [95% limits of agreement] PV_{MPE} vs. PV_{MPI}: −3.73 mL [−11.90; 4.45 mL], *p* < 0.001, paired *t*-test; 95% confidence interval [CI] −4.47; −2.99). The inter-reader agreement between PV_{EF1} and PV_{EF2} is shown in Fig. 4b. The mean difference (95% limits of agreement) between reader 1 and reader 2 was −0.78 mL (−15.08; 13.51 mL). The ICC (95% CI) was 0.93 (0.96; 0.953).

Table 2 and the box and whisker plot in Fig. 5 show the mean, median, minimum, and maximum values for all compared volumes. Supplemental Table 2 in electronic supplementary material shows which of the combinations of the compared volumes statistically significant differences were found. Timed observations for planimetry by an experienced reader were recorded in 14/124 patients, and mean time consumption per case was 8 min 4 s.

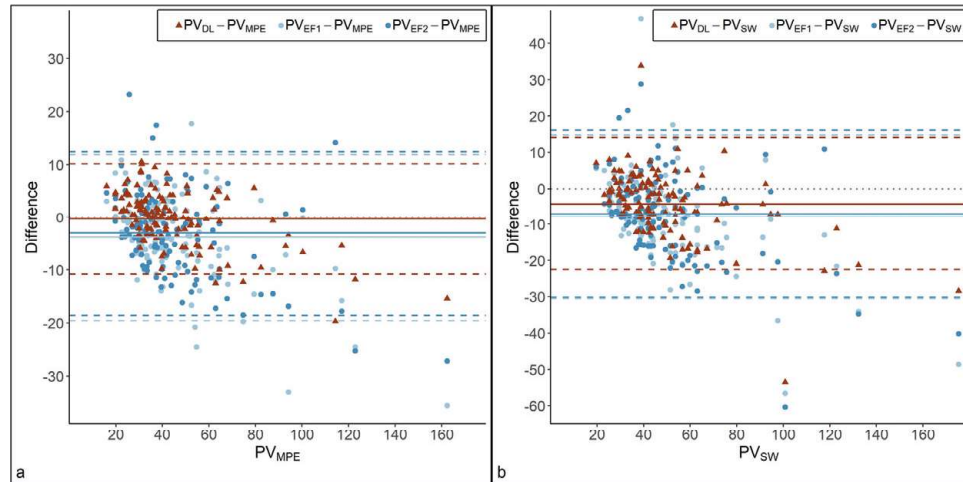


Fig. 3 Bland Altman plot comparing two methods of measuring prostate volume: automated deep learning (PV_{DL}) and manual ellipsoid formula performed by two radiologists (PV_{EF1} and PV_{EF2}). **a** Prostate volumes from manual planimetry by an experienced radiologist (PV_{MPE}) were used as a reference standard. **b** Prostate volumes from prostatectomy

specimen weight (PV_{SW}) were used as a reference standard. The solid lines represent the mean difference between methods and the dashed lines limits of agreement, calculated as the mean difference \pm 1.96SD. The dotted line represents no difference between methods

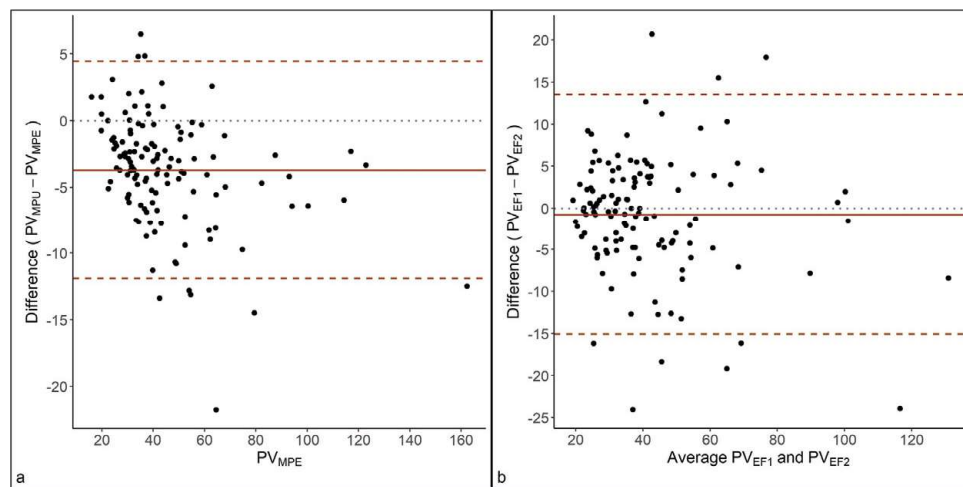


Fig. 4 **a** Bland Altman plot comparing manual planimetry by experienced (PV_{MPE}) and inexperienced radiologists (PV_{MPU}) for measuring prostate volume. **b** Bland Altman plot comparing prostate volumes measured by the ellipsoid formula by two radiologists. The

solid lines represent the mean difference between methods and the dashed lines limits of agreement, calculated as the mean difference \pm 1.96SD. The dotted line represents no difference between methods

Table 2 Eight prostate volume estimates

Measurement	Overall (N = 124)
PV _{EF1}	
Mean (SD)	41.2 (19.4)
Median [min, max]	35.8 [17.2, 127]
PV _{EF2}	
Mean (SD)	42.0 (20.5)
Median [min, max]	36.0 [18.7, 135]
PV _{MPU}	
Mean (SD)	41.3 (21.4)
Median [min, max]	35.5 [17.4, 150]
PV _{MPE}	
Mean (SD)	45.0 (22.5)
Median [min, max]	39.1 [16.0, 162]
PV _{DL}	
Mean (SD)	44.7 (19.9)
Median [min, max]	39.7 [21.4, 147]
PV _{TRUS}	
Mean (SD)	40.2 (20.1)
Median [min, max]	35.5 [17.0, 140]
Missing	2 (1.6%)
PV _{SD}	
Mean (SD)	39.6 (17.8)
Median [min, max]	37.9 [15.6, 153]
PV _{SW}	
Mean (SD)	48.9 (23.1)
Median [min, max]	43.1 [19.5, 175]
Missing	1 (0.8%)

Discussion

This study shows that the PVs obtained from MRI using a commercially available deep learning algorithm have better agreement and precision with two reference standard volumes than the currently recommended gold standard method performed by a radiologist. There was good inter-reader agreement between two experienced radiologists using the ellipsoid formula. Furthermore, this study indicates a small but significant mean difference (low bias) with good precision when evaluating inter-reader agreement between experienced and inexperienced radiologists performing manual planimetry.

In 2013, Turkbey et al [29] showed highly accurate PV estimates by a non-commercially available deep learning algorithm compared to specimen volumes in 99 patients using correlation and Dice similarity coefficients [30] (Pearson coefficient 0.88–0.91, DICE 0.89). They reported no difference compared to manual planimetry. This result is similar in the present study, as the bias between the deep learning algorithm and manual planimetry was close to zero (−0.33 mL). In 2018, Bezinque et al [12] showed in 99 patients that the most reliable method for

PV measurement was manual planimetry by an inexperienced reader (ICC 0.91), followed by ellipsoid formula on MRI (ICC 0.9), compared to manual planimetry by an experienced reader. The authors concluded that automated segmentations (ICC 0.38) made by commercially available software must be individually assessed for accuracy, contradictory to the results of the present study. The amount of training data available for algorithms is rapidly increasing, which may explain the difference in performance in the current study and older studies. In 2020, Lee et al [19] evaluated a non-commercially available deep learning algorithm on a 330 MRI (260 training and 70 test case sets) using manual planimetry as the reference standard. The authors concluded that the algorithm provided reliable PV estimates (ICC 0.90) compared to those obtained with the ellipsoid formula (ICC 0.92). The mean error between algorithm and manual planimetry was 2.5 mL (0.33 mL in our study) and 3.3 mL between ellipsoid formula and manual planimetry (3.05–3.83 mL in our study). In 2021, Salvaggio et al [31] used manual planimetry as the reference standard when they evaluated two deep learning algorithms for prostate segmentation in a cohort of 103 patients. The authors concluded that the presence of median lobe enlargement may lead to overestimation by the ellipsoid formula, recommending a segmentation method. In 2021, Cuocolo et al [32] compared different deep learning algorithms in 204 patients (99 training sets and 105 test sets) using manual planimetry as the reference standard. The efficient neural network (ENet) showed the best performance with the lowest relative volume difference compared to reference standard.

In our study, the mean difference compared to specimen weight volumes (using the specimen gravity formula) was less than zero for both the deep learning algorithm and the two ellipsoid formula measurements (−4.22 mL, −7.89 mL, and −6.97 mL, respectively), indicating a systematic underestimation in line with Bezinque et al [12]. On the other hand, Paterson et al [13] showed that the ellipsoid formula overestimated by a mean 1.4 mL and Lee et al [19] that it overestimated by 2.4%. The discrepant results may be related to a different proportion of cases with median lobe hypertrophy or measurement difficulties and inconsistencies when dealing with extraprostatic tissue.

As described by Salvaggio [31], the correlation between PVs obtained from MRI or radical prostatectomy specimens is dependent on the PV itself, a tendency that we also saw in our material, with overestimation of small prostate gland volumes and underestimation of large prostate gland volumes.

Our study showed good inter-reader agreement between two radiologists estimating PVs with the ellipsoid formula according to PIRADS v 2.1 (mean difference −0.78) and ICCs that were in line with two previous studies [17, 33].

Compared to the present study, Bezinque [12] reported somewhat better agreement between experienced and inexperienced radiologists, with a mean difference of −1.00 mL and ICC of 0.91. Differences in study cohorts and design make

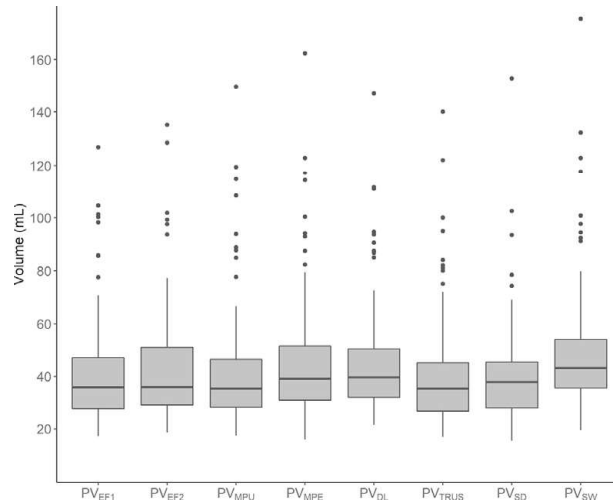


Fig. 5 Distribution of volumes according to eight different methods for measuring prostate volume. PV_{EF1} = prostate volume by ellipsoid formula radiologist 1, PV_{EF2} = prostate volume by ellipsoid formula radiologist 2, PV_{MPU} = prostate volume by manual planimetry by inexperienced radiologist, PV_{MPE} = prostate volume by manual planimetry by experienced radiologist, PV_{DL} = prostate volume by deep

learning, PV_{TRUS} = prostate volume by transrectal ultrasound, PV_{SD} = prostate volume based on specimen dimensions, PV_{SW} = prostate volume based on specimen weight. The center line denotes the median value, the box contains the 25th to 75th percentiles of dataset. The whiskers mark the 5th and 95th percentiles, and values beyond these upper and lower bounds are considered outliers, marked with dots

comparisons between the studies difficult. A small amount of bias can be acceptable as long as the precision is good (as shown by narrow limits of agreement), which seems to be the case with our results.

In this study, we evaluated the agreement between deep learning and ellipsoid formula-determined volumes against two different reference standards. The first evaluation against manual planimetry as the reference standard which was also performed by Cuocolo, Lee, and Bezinque [12, 19, 32]. To avoid the results relying mainly on the similar methodology between deep learning and manual planimetry (i.e., whole gland segmentation by outer contours), we performed a sensitivity analysis by changing the reference standard to the PV based on specimen weight, as used by Turkbey, Mazaheri, and Bulman [17, 23, 29]. The results show that, for both reference standards, the deep learning PVs had lower bias and narrower limits of agreement, meaning better precision than the ellipsoid formula volumes.

To compare the agreement between methods for PV measurement, we based our statistical analysis and visualizations on Bland Altman plots. Several previous studies comparing methods for measuring PV [12, 17, 29] have used different statistical methods based on correlation, and several studies have used DICE similarity coefficients [19, 32]. Correlation coefficients tell us about the linear relationship but not about the agreement, which is indicated by the limits of agreement. In

our opinion, the DICE coefficients do not add value to the comparison of methods as evaluated in this study, but they play a key role when studying the quality of outer contour delineation for MRI/ultrasound fusion biopsy and brachytherapy planning. This is the topic of a planned future study by our research group.

Our study has several strengths. The deep learning algorithm was previously unexposed to the data and, to the best of our knowledge, the test set was larger than in previous studies [12, 17, 19, 29, 32]. To reflect a true clinical context, we used a heterogeneous MRI data set with a multicenter, multi-scanner setup. We used a proprietary commercially available deep learning algorithm, whereas Lee [19] evaluated a non-commercially available 3D deep learning convolutional neural network. Cuocolo [32] compared three different deep learning networks (ENet, ERFNet, and U-net), concluding that deep learning networks can accurately segment the prostate and Enet performed best. In this study, all MRI exams were resampled to isotropic voxel size and to identical matrix resolution to facilitate the deep learning segmentation and, in our study, no resampling was necessary despite variations in MRI protocols.

We investigated the possibility to perform a validation of the deep learning algorithm on publicly available datasets. The available public datasets with manual planimetry as reference standard [34, 35] had been included in early training of the algorithm, why a performance test could not be performed as data sets for training, validation, and testing must always be

unique and separated. It is reasonable to believe this applies also for other commercially available deep learning algorithms, which emphasizes the importance of study designs like the current one, using heterogeneous clinically relevant datasets when testing the robustness of AI models.

This study has some limitations. Firstly, only one algorithm was tested. In addition, although the MRI data set is heterogeneous, one vendor is dominant, which is also the same company behind the evaluated algorithm. Furthermore, the cohort was only prostatectomy patients, which does not reflect the clinical setting, where a larger variation of both cancer and non-cancer patients is scanned. The evaluated algorithm does not offer sub-segmentation of transition and peripheral zone separately. Sub-segmentation would enable more elaborated density calculation, like prostate volume index and transition zone PSAD [36, 37]. Those measurements can add prognostic value for use in a clinical setting with mixed patients (with cancer, no cancer and inflammation). However, in this current study, with cancer cases only, they play a minor role. We plan on future studies dealing with these limitations by evaluating several algorithms with different scanner vendors and a more heterogeneous patient group. A future study should ideally be designed as a non-inferiority study with power calculation for adequate cohort size. Only one experienced radiologist performed manual planimetry. However, it is known there is an interreader variability in manual planimetry [38]. We tried to take this into consideration by letting a second radiologist (less experienced) also perform manual planimetry and via performing the sensitivity analysis with specimen volumes.

Clinical implications

The number of U.S. FDA-approved commercial AI/deep learning algorithms is rapidly increasing [21], but there is a concern regarding the challenges accompanying the application of those algorithms in the clinical routine. There is a need for structured monitoring and follow-up when we start using those algorithms in day-to-day practice.

To the best of our knowledge, no earlier studies have evaluated a proprietary commercially available deep learning algorithm on such a heterogeneous MRI data set as in this study. The current study setup with a multicenter, multi-scanner, multi-parameter protocol resembles the true clinical situation in a diversified national or international setting.

Conclusion

A deep learning algorithm is at least as good as the PI-RADS 2.1-based ellipsoid formula for assessing PV when compared to in vivo and ex vivo reference standards. This is a promising

step towards algorithms helping reallocate radiologist resources towards more complex work tasks than manually measuring PVs.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00330-022-09239-8>.

Acknowledgements We gratefully thank secretary Kajsa Trens at the department of translational medicine for helping out during data collection and statistician Andrea Dahl Sturedahl at Forum Söder for statistical support.

Funding Open access funding provided by Lund University. This study has received funding by grants from Governmental funding for clinical research (ALF), grants for PhD students from Region Skåne and scholarship from Stig and Ragna Gorthon Foundation.

Declarations

Guarantor The scientific guarantor of this publication is Professor Sophia Zackrisson.

Conflict of interest The authors of this manuscript declare relationships with the following companies:

Sophia Zackrisson has received speaker fees from Siemens Healthcare AB and Pfizer AB. Anders Bjartell has received grants/research supports from Ferring, Bayer and Merck. AB has received honoraria or consultation fees from Astellas, AstraZeneca, Bayer, Janssen, Merck, Recordati and Sandoz. AB has participated in company sponsored speaker's bureau by Astellas, Bayer, IPSEN, Janssen, Recordati and Sandoz. AB is stock shareholder in LIDDS Pharma, Glactone Pharma and WntResearch.

All other authors declare no conflicts of interest.

Statistics and biometry Andrea Dahl Sturedahl kindly provided statistical advice for this manuscript.

No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology

- retrospective
- diagnostic or prognostic study
- multicenter study

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a

credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Garvey B, Türkbey B, Truong H et al (2014) Clinical value of prostate segmentation and volume determination on MRI in benign prostatic hyperplasia. *Diagn Interv Radiol* 20:229–233
- Heidler S, Drerup M, Lusuadi L et al (2018) The correlation of prostate volume and prostate-specific antigen levels with positive bacterial prostate tissue cultures. *Urology* 115:151–156
- Kim YM, Park S, Kim J et al (2013) Role of prostate volume in the early detection of prostate cancer in a cohort with slowly increasing prostate specific antigen. *Yonsei Med J* 54:1202–1206
- Sim KC, Sung DJ, Kang KW et al (2017) Magnetic resonance imaging-based prostate-specific antigen density for prediction of Gleason Score upgrade in patients with low-risk prostate cancer on initial biopsy. *J Comput Assist Tomogr* 41:731–736
- Nordström T, Akre O, Aly M et al (2018) Prostate-specific antigen (PSA) density in the diagnostic algorithm of prostate cancer. *Prostate Cancer Prostatic Dis* 21:57–63
- Fascelli M, Rais-Bahrami S, Sankineni S et al (2016) Combined biparametric prostate magnetic resonance imaging and prostate-specific antigen in the detection of prostate cancer: a validation study in a biopsy-naïve patient population. *Urology* 88:125–134
- Loeb S, Bruinsma SM, Nicholson J et al (2015) Active surveillance for prostate cancer: a systematic review of clinicopathologic variables and biomarkers for risk stratification. *Eur Urol* 67:619–626
- Blackwell KL, Bostwick DG, Myers RP et al (1994) Combining prostate specific antigen with cancer and gland volume to predict more reliably pathological stage: the influence of prostate specific antigen cancer density. *J Urol* 151:1565–1570
- Ahmed HU, El-Shater Bosaily A et al (2017) Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet* 389: 815–822
- Kasivisvanathan V, Rannikko AS, Borghi M et al (2018) MRI-targeted or standard biopsy for prostate-cancer diagnosis. *N Engl J Med* 378:1767–1777
- Türkbey B, Rosenkrantz AB, Haider MA et al (2019) Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2. *Eur Urol* 76:340–351
- Bezinque A, Moriarty A, Farrell C et al (2018) Determination of prostate volume: a comparison of contemporary methods. *Acad Radiol* 25:1582–1587
- Paterson NR, Lavallée LT, Nguyen LN et al (2016) Prostate volume estimations using magnetic resonance imaging and transrectal ultrasound compared to radical prostatectomy specimens. *Can Urol Assoc J* 10:264
- Karademir I, Shen D, Peng Y et al (2013) Prostate volumes derived from MRI and volume-adjusted serum prostate-specific antigen: correlation with Gleason score of prostate cancer. *Am J Roentgenol* 201:1041–1048
- Cheng R, Lay NS, Roth HR et al (2019) Fully automated prostate whole gland and central gland segmentation on MRI using holistically nested networks with short connections. *J Med Imaging* 6: 024007
- Jeong CW, Park HK, Hong SK et al (2008) Comparison of prostate volume measured by transrectal ultrasonography and MRI with the actual prostate volume measured after radical prostatectomy. *Urol Int* 81:179–185
- Bulman JC, Toth R, Patel AD et al (2012) Automated computer-derived prostate volumes from MR imaging data: comparison with radiologist-derived MR imaging and pathologic specimen volumes. *Radiology* 262:144–151
- Cuocolo R, Cipullo MB, Stanzione A et al (2019) Machine learning applications in prostate cancer magnetic resonance imaging. *Eur Radiol Exp* 3:1–8
- Lee DK, Sung DJ, Kim CS et al (2020) Three-dimensional convolutional neural network for prostate MRI segmentation and comparison of prostate volume measurements by use of artificial neural network and ellipsoid formula. *AJR Am J Roentgenol* 214: 1229–1238
- Ma L, Guo R, Zhang G et al (2017) Automatic segmentation of the prostate on CT images using deep learning and multi-atlas fusion. *Medical Imaging 2017: image processing*. International Society for Optics and Photonics, p 1013320
- Allen B, Dreyer K, Stibolt R Jr et al (2021) Evaluation and real-world performance monitoring of artificial intelligence models in clinical practice purchase: try it, buy it, check it. *J Am Coll Radiol*. <https://doi.org/10.1016/j.jacr.2021.08.022>
- Winkel DJ, Heye T, Weikert TJ et al (2019) Evaluation of an AI-based detection software for acute findings in abdominal computed tomography scans: toward an automated work list prioritization of routine CT examinations. *Invest Radiol* 54:55–59
- Mazaheri Y, Goldman DA, Di Paolo PL et al (2015) Comparison of prostate volume measured by endorectal coil MRI to prostate specimen volume and mass after radical prostatectomy. *Acad Radiol* 22: 556–562
- Ohlén S, Ekman P, Ringertz H (1982) Assessment of prostatic size with computed tomography. Methodologic aspects. *Acta Radiol Diagn (Stockh)* 23:219–223
- Varna M, Morgan JM (2010) The weight of the prostate gland is an excellent surrogate for gland volume. *Histopathology* 57:55–58
- Yang D, Xu D, Zhou SK et al (2017) Automatic liver segmentation using an adversarial image-to-image network. *International conference on medical image computing and computer-assisted intervention*. Springer, pp 507–515
- Egevad L, Srigley JR, Delahunt B (2011) International society of urological pathology consensus conference on handling and staging of radical prostatectomy specimens. *Adv Anat Pathol* 18:301–305
- Team RC (2020) R: a language and environment for statistical computing. Version 4.0. 2. Vienna, Austria
- Türkbey B, Fotin SV, Huang RJ et al (2013) Fully automated prostate segmentation on MRI: comparison with manual segmentation methods and specimen volumes. *Am J Roentgenol* 201:W720–W729
- Dice LR (1945) Measures of the amount of ecologic association between species. *Ecology* 26:297–302
- Salvaggio G, Comelli A, Portoghesi M et al (2021) Deep learning network for segmentation of the prostate gland with median lobe enlargement in T2-weighted MR images: comparison with manual segmentation method. *Curr Probl Diagn Radiol*. <https://doi.org/10.1067/j.cpradiol.2021.06.006>
- Cuocolo R, Comelli A, Stefano A et al (2021) Deep learning whole-gland and zonal prostate segmentation on a public MRI dataset. *J Magn Reson Imaging* 54:452–459
- Ghafoor S, Becker AS, Woo S et al (2020) Comparison of PI-RADS Versions 2.0 and 2.1 for MRI-based calculation of the

- prostate volume. *Acad Radiol*. <https://doi.org/10.1016/j.acra.2020.07.027>
34. Litjens G, Toth R, van de Ven W et al (2014) Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge. *Med Image Anal* 18:359–373
 35. Armato SG 3rd, Huisman H, Drukker K et al (2018) PROSTATEx Challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images. *J Med Imaging (Bellingham)* 5:044501
 36. Porcaro AB, Tafuri A, Sebben M et al (2019) Prostate volume index is able to differentiate between prostatic chronic inflammation and prostate cancer in patients with normal digital rectal examination and prostate-specific antigen values <10 ng/mL: results of 564 Biopsy Naïve Cases. *Urol Int* 103:415–422
 37. Schneider AF, Stocker D, Hötter AM et al (2019) Comparison of PSA-density of the transition zone and whole gland for risk stratification of men with suspected prostate cancer: a retrospective MRI-cohort study. *Eur J Radiol* 120:108660
 38. Becker AS, Chaitanya K, Schawkat K et al (2019) Variability of manual segmentation of the prostate in axial T2-weighted MRI: a multi-reader study. *Eur J Radiol* 121:108716

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Paper II



Paper III





OPEN ACCESS

EDITED BY
Jasper Nijkamp,
Aarhus University, Denmark

REVIEWED BY
Petra J. Van Houdt,
The Netherlands Cancer Institute (NKI),
Netherlands
Catarina Fernandes,
Eindhoven University of Technology,
Netherlands

*CORRESPONDENCE
Johan Bengtsson
✉ johan.bengtsson@med.lu.se

SPECIALTY SECTION
This article was submitted to
Cancer Imaging and
Image-directed Interventions,
a section of the journal
Frontiers in Oncology

RECEIVED 24 October 2022
ACCEPTED 03 February 2023
PUBLISHED 20 February 2023

CITATION
Bengtsson J, Thimansson E, Baubeta E,
Zackrisson S, Sundgren PC, Bjartell A and
Flondell-Sité D (2023) Correlation between
ADC, ADC ratio, and Gleason Grade
group in prostate cancer patients
undergoing radical prostatectomy:
Retrospective multicenter study
with different MRI scanners.
Front. Oncol. 13:1079040.
doi: 10.3389/fonc.2023.1079040

COPYRIGHT
© 2023 Bengtsson, Thimansson, Baubeta,
Zackrisson, Sundgren, Bjartell and Flondell-
Sité. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Correlation between ADC, ADC ratio, and Gleason Grade group in prostate cancer patients undergoing radical prostatectomy: Retrospective multicenter study with different MRI scanners

Johan Bengtsson^{1,2*}, Erik Thimansson^{3,4}, Erik Baubeta^{2,3},
Sophia Zackrisson^{3,5}, Pia Charlotte Sundgren^{1,2,6},
Anders Bjartell^{3,7} and Despina Flondell-Sité^{3,7}

¹Department of Clinical Sciences, Radiology, Lund, Lund University, Lund, Sweden, ²Department of Medical Imaging and Physiology, Skåne University Hospital, Lund, Sweden, ³Department of Translational Medicine, Lund University, Malmö, Sweden, ⁴Department of Radiology, Helsingborg Hospital, Helsingborg, Sweden, ⁵Department of Medical Imaging and Physiology, Skåne University Hospital, Malmö, Sweden, ⁶Lund Bioimaging Center (LBIC), Lund University, Lund, Sweden, ⁷Department of Urology, Skåne University Hospital, Malmö, Sweden

Background: MRI is an important tool in the prostate cancer work-up, with special emphasis on the ADC sequence. This study aimed to investigate the correlation between ADC and ADC ratio compared to tumor aggressiveness determined by a histopathological examination after radical prostatectomy.

Methods: Ninety-eight patients with prostate cancer underwent MRI at five different hospitals prior to radical prostatectomy. Images were retrospectively analyzed individually by two radiologists. The ADC of the index lesion and reference tissues (contralateral normal prostatic, normal peripheral zone, and urine) was recorded. Absolute ADC and different ADC ratios were compared to tumor aggressiveness according to the ISUP Gleason Grade Groups extracted from the pathology report using Spearman's rank correlation coefficient (ρ). ROC curves were used to evaluate the ability to discriminate between ISUP 1-2 and ISUP 3-5 and intra class correlation and Bland-Altman plots for interrater reliability.

Results: All patients had prostate cancer classified as ISUP grade ≥ 2 . No correlation was found between ADC and ISUP grade. We found no benefit of using the ADC ratio over absolute ADC. The AUC for all metrics was close to 0.5, and no threshold could be extracted for prediction of tumor aggressiveness. The interrater reliability was substantial to almost perfect for all variables analyzed.

Conclusions: ADC and ADC ratio did not correlate with tumor aggressiveness defined by ISUP grade in this multicenter MRI study. The result of this study is opposite to previous research in the field.

KEYWORDS

MRI, MR-diffusion, ADC, neoplasms, prostate

Introduction

Prostate cancer (PCa) is the most common cancer in men worldwide (GLOBOCAN 2020) (1). However, most men with PCa have low-grade, indolent tumors. Therefore, discriminating between indolent and aggressive tumors is a diagnostic issue. With the traditional diagnostic approach, which includes a blood test of prostate specific antigen (PSA), digital rectal examination, and systematic transrectal ultrasound-guided biopsies, only a small and randomly distributed fraction of the gland is examined, resulting in a substantial risk of both over- and under-sampling. A more modern pathway involves magnetic resonance imaging (MRI) to detect clinically significant prostate cancer (csPCa) and rule out other causes of elevated PSA levels. On pathology, csPCa is defined as a Gleason score ≥ 7 (including 3 + 4 with a prominent but not predominant Gleason 4 component), volume ≥ 0.5 mL, and/or extra prostatic extension (2). Today, the International Society of Urological Pathology (ISUP) grade is often used to categorize different Gleason score patterns (3). When using MRI as a triage tool, unnecessary biopsies can be avoided, and targeted when required. This approach was investigated in the PRECISION study, which showed that MRI followed by targeted biopsies detected more significant tumors (38% versus 26%, $p=0.005$) and fewer insignificant tumors (9% versus 22%, $p<0.001$) compared to systematic biopsies (4). In the group that had an MRI in the work up, 28% had a negative MRI and, thus, did not have to undergo biopsy. These results changed the work-up routine, and MRI is now a cornerstone of PCa diagnosis. Therefore, the demands on MRI are high in terms of technical quality and radiological interpretation for correctly detecting or excluding csPCa.

Prostate Imaging – Reporting and Data System (PI-RADS, version 2.1) is a system that describes how to perform, interpret, and report MRI of the prostate (2). The most important MRI sequence is diffusion-weighted imaging (DWI), which is the deciding sequence in the peripheral zone (PZ) and the secondary sequence in the transition zone (TZ). DWI provides information on tissue composition and tumor cellularity (5). The signal intensity on DWI reflects the motion of water molecules in the tissue. The concept is based on the theory that a tumor consists of more dense tissue than normal prostatic tissue.

Several studies have shown that the ADC value inversely correlates with ISUP grade and is often used as a marker of aggressiveness (5–9). Several cut-off values have been proposed; in PIRADS 2.0, a threshold of 750–900 $\mu\text{m}^2/\text{s}$ was suggested as a pathological ADC value, but no consensus has been reached (8, 10).

The concept is associated with several difficulties. First, the ADC varies substantially depending on several factors, including the b-values used, scanner field strength, patient and coil geometry, temporal fluctuations in the magnet, and variations in measurements between readers. Furthermore, non-cancerous lesions, such as benign prostate hypertrophy, may also exhibit decreased ADC values, and there is a substantial overlap in ADC values and PCa (11). ADC is sometimes used as a marker of aggressiveness in other organs and diseases. For example, in rectal adenocarcinoma, a lower ADC value is associated with a more aggressive tumor and poorer survival rate. Similar correlations have been found in certain types of breast cancer, ovarian cancer, lung cancer, and gliomas (12–15).

A common way to overcome the differences in absolute ADC values is to normalize the ADC by using different ADC ratios (10, 16). The ADC ratio is expressed as the ratio between the ADC value of the tumor and the ADC value of another location, such as non-cancerous tissue in the same organ or other organs in the same patient (5, 17).

In recent years, several studies have investigated the potential benefit of using the ADC ratio over absolute ADC values. Some authors have affirmed that the ADC ratio is the preferred method and demonstrated significant capability in discriminating Gleason 3 + 4 from 4 + 3 PCa (5, 8, 9, 16, 18). Other authors have been more doubtful (19).

The aim of the present study was to investigate, in a consecutive patient cohort imaged using different MRI scanners, how absolute ADC value and ADC ratios correlate with ISUP grade following robot-assisted laparoscopic prostatectomy (RALP). A secondary aim was to assess the potential inter-observer variability.

Material and methods

The study was a retrospective cohort study approved by the local ethics review committee at Lund University (Dnr 2014-886) and the Swedish ethical review authority (entry no. 2019-03674).

Study population

All consecutive patients who underwent RALP for biopsy proven PCa at Skåne University Hospital in Malmö, Sweden, during 2018 were identified and assessed for eligibility. Patients

were included if they had undergone MRI less than 1 year before surgery at five different hospitals. Patients were excluded if the index lesion described in the pathology report was not identified on MRI, severe artifacts were present on MRI, the MRI was performed outside of Region Skåne, or the patient opted out. Lesions were excluded based on consensus between two readers (JB and ET). The data collection algorithm is presented in Figure 1. Patient data were obtained from medical records.

Pathological examination

The surgical specimens were handled according to clinical routines and fixed in formalin. Lesions were examined by experienced pathologists using hematoxylin and eosin staining. Pathological data and whole mount (WM) tumor maps were obtained from the pathology report. The location and Gleason score of the index lesion were recorded using the ISUP category classification (Table 1).

MRI acquisition and image analysis

Preoperative MRI of the prostate was performed within Region Skåne using one of eight MRI scanners at five sites. Both 3T and 1.5T scanners were used. According to local routines, different imaging acquisition parameters were used at different sites. All protocols included transverse, coronal, and sagittal T2-weighted turbo spin-echo images, transverse T1-weighted images, diffusion-weighted images with a high b-value of 1500 s/mm², and a calculated ADC map. A list of MRI scanners and imaging acquisition parameters for the DWI are presented in Table 2.

Two readers, both specialists in radiology with 4 and 5 years of experience in reading prostate MRI, performed all imaging analyses as described below. The examinations were reviewed using the clinical Picture Archiving and Communication System, Sectra IDS7.

First, and in consensus, the two readers matched the index lesion in the surgical specimen with the corresponding lesion on MRI using the

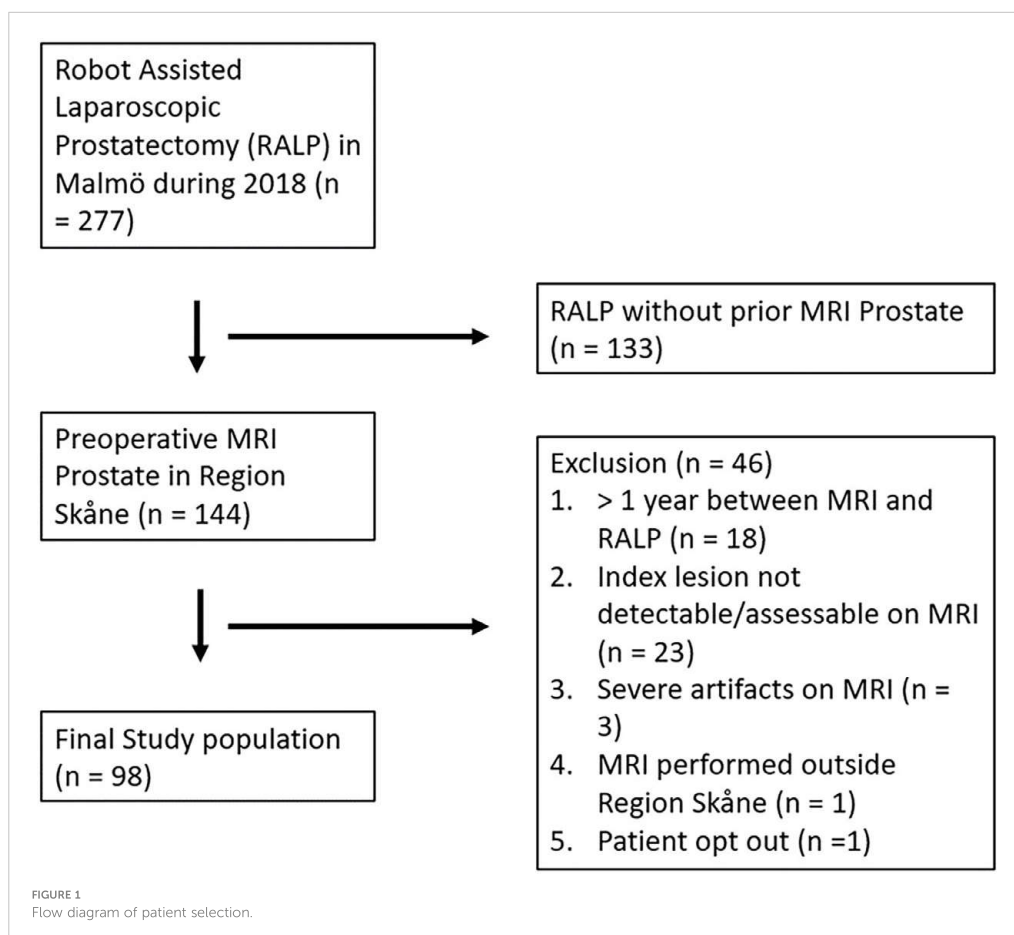


TABLE 1 ISUP grade groups and the corresponding Gleason scores and patterns.

ISUP grade group	Gleason score	Gleason pattern
1	≤6	≤3+3
2	7	3+4
3	7	4+3
4	8	4+4, 3 + 5, 5 + 3
5	9 or 10	4+5, 5 + 4, or 5 + 5

pathological report and the whole-mount tumor map. In a second step, the remaining interpretation and image analyses were performed individually. For the index lesion, each reader recorded the maximum diameter in millimeters, zone location (PZ or TZ), and PI-RADS score (version 2.1). A circular region of interest (ROI) was placed in the index lesion in the ADC map on the slice with the largest cross-sectional area of tumor (ADC_{lesion}). The ROI was drawn to include only the lesion without any surrounding parenchyma. The size of the ROI was not fixed, it was drawn as big as possible within the defined lesion. A second ROI ($ADC_{\text{contralat ref}}$) of the same size was placed at the contralateral position on the same slice, that is in the same zone as the index lesion. A third and fourth ROI was placed in the most homogenous area in the PZ ($ADC_{\text{PZ ref}}$) and in the urinary bladder ($ADC_{\text{urine ref}}$), respectively. For each ROI, the mean ADC value was recorded (Figure 2).

Statistical analysis

Descriptive statistics were used to present the study population. Box plots and Spearman's rank correlation coefficient (ρ) were used to evaluate the association between ISUP grade and ADC variables. Measurements from reader 1 were used for the analyses of ADC metrics. These analyses were repeated and stratified by scanner field strength (1.5 vs. 3T) and tumor location (PZ vs. TZ). Receiver operating characteristic (ROC) curves were used to evaluate the ability to discriminate between ISUP 1-2 and ISUP 3-5 based on ADC variables. Interrater reliability was evaluated using Bland-Altman plots and intraclass correlation (ICC) based on the formula for random effects, absolute agreement, and single rater measurements. The ICC values were rated as follows: slight agreement, 0 – 0.20; fair agreement, 0.21 – 0.40; moderate agreement, 0.41 – 0.60; substantial agreement, 0.61 – 0.80; almost perfect agreement, 0.81 – 1.

All statistical analyses were performed in R version 4.0.2. The pROC package was used for ROC curves and the irr package to calculate ICC.

Results

A total of 144 men underwent RALP due to biopsy proven PCa and had an MRI prior to the procedure. After exclusion for different reasons (Figure 1), 98 patients were included in the final study analysis. The patient and tumor characteristics are presented in

Table 3. No specimen was classified as ISUP 1. Most index lesions were located in the PZ of the prostate. Patients with different ISUP grades were relatively evenly distributed over the eight scanners, details are available in Supplementary Table 1.

ADC measurements vs. ISUP grade

The average ADC_{lesion} was $652 \times 10^{-6} \text{ mm}^2/\text{s}$ (range $396 \times 10^{-6} \text{ mm}^2/\text{s}$ to $1271 \times 10^{-6} \text{ mm}^2/\text{s}$), whereas the average $ADC_{\text{contralat ref}}$ tissue was $1275 \times 10^{-6} \text{ mm}^2/\text{s}$ (range $779 \times 10^{-6} \text{ mm}^2/\text{s}$ to $1794 \times 10^{-6} \text{ mm}^2/\text{s}$). The average $ADC_{\text{PZ ref}}$ was $1478 \times 10^{-6} \text{ mm}^2/\text{s}$ (range $779 \times 10^{-6} \text{ mm}^2/\text{s}$ to $2155 \times 10^{-6} \text{ mm}^2/\text{s}$) and of $ADC_{\text{urine ref}}$ was $2021 \times 10^{-6} \text{ mm}^2/\text{s}$ (range $861 \times 10^{-6} \text{ mm}^2/\text{s}$ to $3368 \times 10^{-6} \text{ mm}^2/\text{s}$; Figure 3). We found no significant negative correlation, between absolute the ADC value of the index lesion and the ISUP grade. The observed spearman correlation between the ADC of the index lesion and ISUP grade was low ($\rho = -0.18$) and not significant. Furthermore, the ADC of the index lesion did not perform well in discriminating between ISUP 1-2 and ISUP 3-5 (AUC = 0.62 [95% CI 0.51-0.74]). A tendency for a negative correlation was observed when the results from the 3T scanners were analyzed separately ($\rho = -0.27$; $p < 0.05$), but not for the 1.5 T scanners ($\rho = -0.01$). Tables reporting the correlation values stratified by field strength are available in Supplementary Table 2. We found no correlation in separate analyses of the PZ and TZ.

The three different ADC ratios were calculated for each lesion ($ADC_{\text{lesion}}/ADC_{\text{contralat ref}}$, $ADC_{\text{lesion}}/ADC_{\text{urine ref}}$ and $ADC_{\text{lesion}}/ADC_{\text{PZ ref}}$) in relation to tumor aggressiveness. None of them showed any discriminatory effect (Figures 3, 4).

The agreement between the two readers in the ADC measurements was almost perfect for ADC_{lesion} (ICC of 0.80 [95% CI 0.72 – 0.86]), $ADC_{\text{contralat ref}}$ (ICC of 0.82 [95% CI 0.75 – 0.88]), and $ADC_{\text{urine ref}}$ (ICC of 0.96 [95% CI 0.94 – 0.97]). For $ADC_{\text{PZ ref}}$ the agreement was substantial (ICC of 0.75 [95% CI 0.65 – 0.86], Figure 5).

Discussion

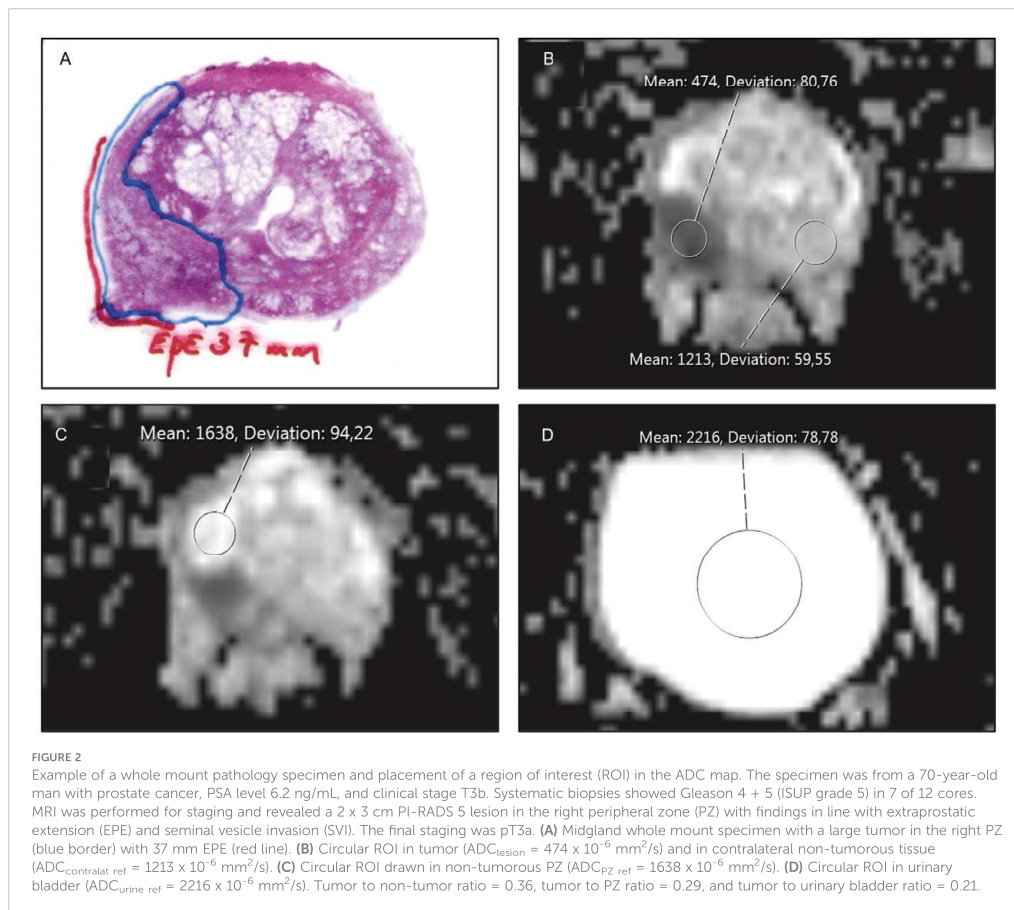
This multi-scanner cohort study of 98 consecutive patients with MRI of the prostate before RALP showed no correlation between the absolute ADC value of the tumor and tumor aggressivity determined by pathology. No improvement was noted when the ADC value was normalized by applying different ADC ratios. Thus, no threshold values for ADC or ADC ratio were determined to discriminate significant from non-significant PCa. The inter-reader agreement between the two observers was substantial to almost perfect.

Different methods of interpretation have been applied to predict whether a lesion found on MRI represents benign tissue, non-significant cancer, or significant cancer. When comparing the results from the different studies, the definition of csPCa is crucial, as most authors try to define a threshold value for different ADC metrics in relation to tumor aggressivity. Some

TABLE 2 Overview of scanners and diffusion-weighted imaging (DWI) acquisition parameters.

Scanner	Patients, n (%)	Vendor	Model	Field strength, T	Sequence	B-values*, s/mm ²	TR, ms	TE, ms	Acquisition matrix	FOV, mm ²	Slice thickness, mm	Inter-slice gap, mm	DWI acquisition time, min
1	36 (36.7)	Siemens	TrioTim	3	2D EPI	50, 400, 1500	5200	87	128	240 x 240	3	3.6	Not applicable
2	12 (12.2)	Siemens	Skyra	3	2D EPI	0, 800, 1500	6700	75	130	240 x 240	3	4	5:50
3	12 (12.2)	Siemens	Prisma	3	2D EPI	0, 800, 1500	4500	66	130	240 x 240	3	3	3:56
4	24 (24.5)	Siemens	Avanto fit	1.5	2D EPI	50, 300, 1000, 1500	3300	67	128	240 x 240	3	4	8:15 + 5:33**
5	7 (7.1)	Siemens	Avanto fit	1.5	2D EPI	50, 300, 1000	2900	62	128	240 x 240	3	3	5:33 + 2:37**
6	3 (3.1)	GE	Optima450r	1.5	2D EPI	50, 400, 800	4746	72	256	280 x 280	4	4.5	3:52
7	2 (2.0)	Siemens	Aera	1.5	Resolve	50, 400, 800	4620	57	116	200 x 200	4	4	6:25 + 6:19**
8	2 (2.0)	Siemens	Avanto fit	1.5	Resolve	50, 400, 800	4620	58	116	200 x 200	4	4	6:41 + 4:41**

EPI, echo-planar imaging; TR, repetition time; TE, echo time; FOV, field of view. *B-values included in the ADC calculation, **separate b1500.



papers have used ISUP grade 1 as non-significant and ISUP 2 and higher as significant (20–24), whereas others have included ISUP 2 in the non-significant group. One study even included all ISUP 2 and 3 in the more harmless group and used the terms intermediate and high-risk cancer as the border between the two groups (25). Boesen et al. performed their analyses on two different cut-offs with ISUP 2 in both the significant and non-significant groups (8). In our study, all resected prostates were ISUP 2 or higher, which gave us no choice to use only ISUP 1 in the non-significant group. This was also true for the 23 patients in whom the index lesion could not be identified on MRI.

Regardless of which definition of csPCa is used, several authors have reported a strong inverse correlation between ADC metrics and tumor aggressivity, with a reported AUC of up to 0.94 (26) or 0.96 (17). This contrasts with the results of our study, as we found an AUC of 0.62, which would suggest that the absolute ADC value is not useful for predicting the presence of csPCa. The reasons for these results can be debated. We used eight different MRI scanners with different acquisition parameters. Disparate

absolute ADC values are not unexpected with these settings. Barret et al. calculated different ADC values from the same scans by combining four b-values in different ways, thereby simulating different parameters (5). Most combinations showed a relatively good inverse correlation with tumor aggressivity. When they used the ratio between tumorous and non-tumorous ADC values, the differences in acquisition parameters were less obvious. Thus, they stated that the ADC ratio may be considered a more robust tool for assessing restricted diffusion in the prostate (5). With the same intention, we evaluated whether the disparate ADC values between our scanners could be more useful when different ratios were applied. However, despite using three different tissues as denominators in the creation of the ratios, no added value or better performance were found for the metrics. In fact, the AUC was even smaller, close to 0.5 for all three ratios, which is slightly smaller than for the absolute ADC. For the 1.5T scanners there was a tendency of positive correlation, instead of the expected negative correlation, between ADC ratio and ISUP grade.

TABLE 3 Patient characteristics (n=98).

Characteristic	Mean \pm SD (min – max)
Age, years	66.3 \pm 6.4 (45 – 76)
Time between MRI and RALP, months	4.08 \pm 2.6 (1 – 11)
Preoperative PSA, ng/mL	9.26 \pm 6.8 (1.8 – 39.0)
	n (%)
Clinical T-stage	
T0	4 (4.1)
T1	13 (13.3)
T1c	29 (29.6)
T2	36 (36.7)
T2b	4 (4.1)
T2c	2 (2.0)
T3	9 (9.2)
T3a	1 (1.0)
Pathological T-stage	
T1	0 (0)
T2	51 (52.0)
T3a	34 (34.7)
T3b	12 (12.2)
T3	0 (0)
Missing	1 (1.0)
Biopsy ISUP grade	
1	7 (7.1)
2	41 (41.8)
3	24 (24.5)
4	9 (9.2)
5	17 (17.3)
Pathological ISUP grade	
1	0 (0)
2	39 (39.8)
3	41 (41.8)
4	3 (3.1)
5	15 (15.3)
MRI field strength	
1.5 Tesla	38 (38.8)
3 Tesla	60 (60.1)
Zone location	
Peripheral zone	68 (69.4)
Transitional zone	30 (30.6)

Several other authors have claimed that the ratio, often tumor versus the contralateral normal appearing tissue, is better than the absolute ADC value. Lebovici et al. showed the usefulness of an ADC ratio in differentiating low-grade and high-grade disease (25). Similar results were reported by Boesen et al. and Litjens et al. (8, 27). Interestingly, both absolute ADC values and the ADC ratios differed considerably between these studies. Itatani et al. assessed 58 men who underwent RALP after MRI and used the internal obturator muscle as the ADC reference, finding superior use of the ratio (AUC 0.85 vs. 0.71) (28). Bajgiran et al. concluded that the ADC ratio is a more robust biomarker of PCa aggressiveness (21). Conversely, Rosencrantz et al. found no benefit of using the ADC ratios with urine ADC as the denominator for differentiating benign and malignant tissue in the PZ (17). Woo et al. (20) included 165 men, and DeCobelli 72 men (26), with contralateral prostatic tissue as the reference and found no benefit of the ADC ratio compared to standalone ADC.

Woo et al. pointed out several reasons why the use of the ADC value for internal reference organs may not yield helpful ADC ratios and thereby add, rather than reduce, sources of error in the interpretation (20). For example, they emphasize that the ADC value of the non-tumor PZ can vary according to age, and that the intrinsically organized chaos of the TZ results in a wide range of normal ADC values (29). Moreover, post-biopsy changes can alter the signal intensity of DWI in the prostatic tissue for several weeks. Finally, as hypothesized by DeCobelli, non-tumorous tissue can be affected by nearby non-visible tumor infiltration or by peritumoral fibrosis and inflammation, which all affect the ADC (26). The b-values that were used to estimate the ADC (Table 2) varied across MRI systems and sites, and several were inconsistent with PI-RADS recommendations (2). For example, the estimation of ADC based on data acquired at low b-values (<100 s/mm²) may introduce a positive bias due to incoherent blood perfusion (30). Furthermore, when the ADC is based on high b-values (>1000 s/mm²), the estimation in normal tissue may be negatively biased due to the rectified noise floor (31). These factors may explain why the ratio did not show a better inverse correlation with cancer aggressiveness than standalone ADC. Moreover, in a systematic review of 39 papers with 2457 patients, Surov et al. identified only a moderate correlation between ADC and Gleason score in PCa located in the PZ, and an even worse correlation in the TZ (32).

Harmonizing MRI parameters between centers is important, especially since the ADC values are used for deciding PI-RADS category and hence, affects the clinical decision. In 2007, the Radiological Society of North America organized The Quantitative Imaging Biomarkers Alliance® (QIBA). QIBA strives for standardization of image acquisition and assesses whether imaging metrics have clinical value (33). Their ongoing work includes evaluation and standardization of DWI in for example MRI Prostate.

In our study, the interrater agreements for different ADC metrics were strong, suggesting that factors other than differences in radiologists' measurements are the reason for the lack of correlation with pathology. Our results are in line with similar previous studies (19, 23, 34).

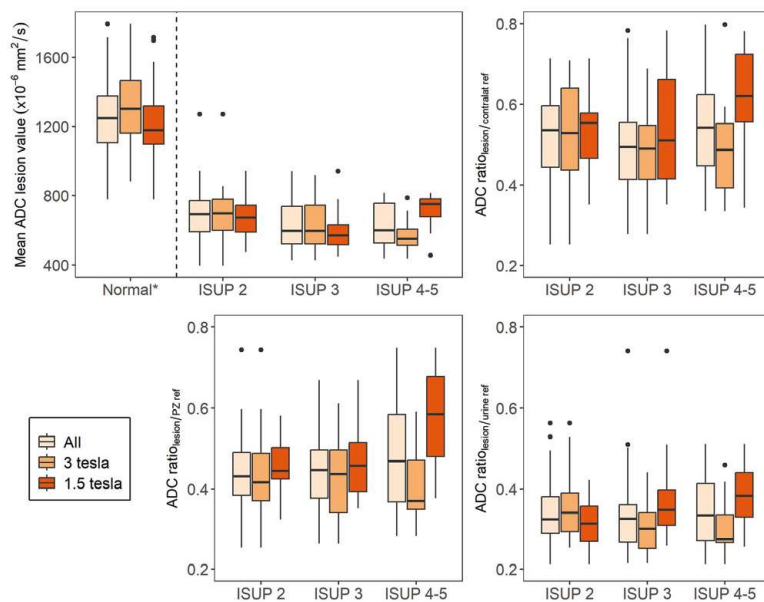


FIGURE 3
Box-and-whisker plots of apparent diffusion coefficient (ADC) metrics for tumors stratified by ISUP grade. (*) *Normal* represents the absolute ADC value of the normal appearing tissue in the contralateral position of the index lesion.

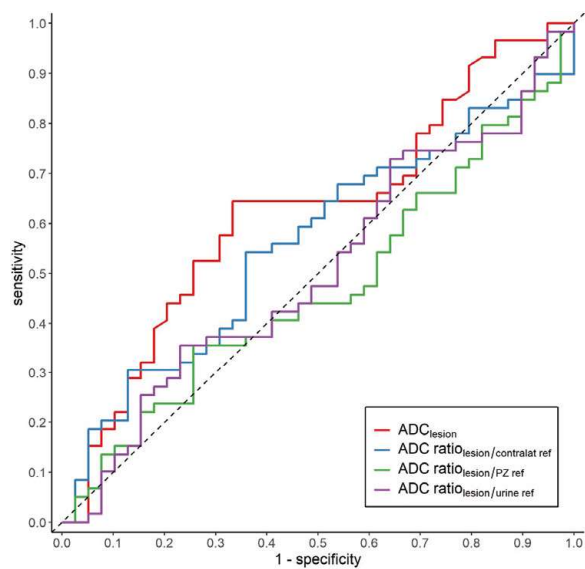


FIGURE 4
ROC curves comparing absolute ADC and three different ADC ratios in discriminating ISUP 1-2 from ISUP 3-5.

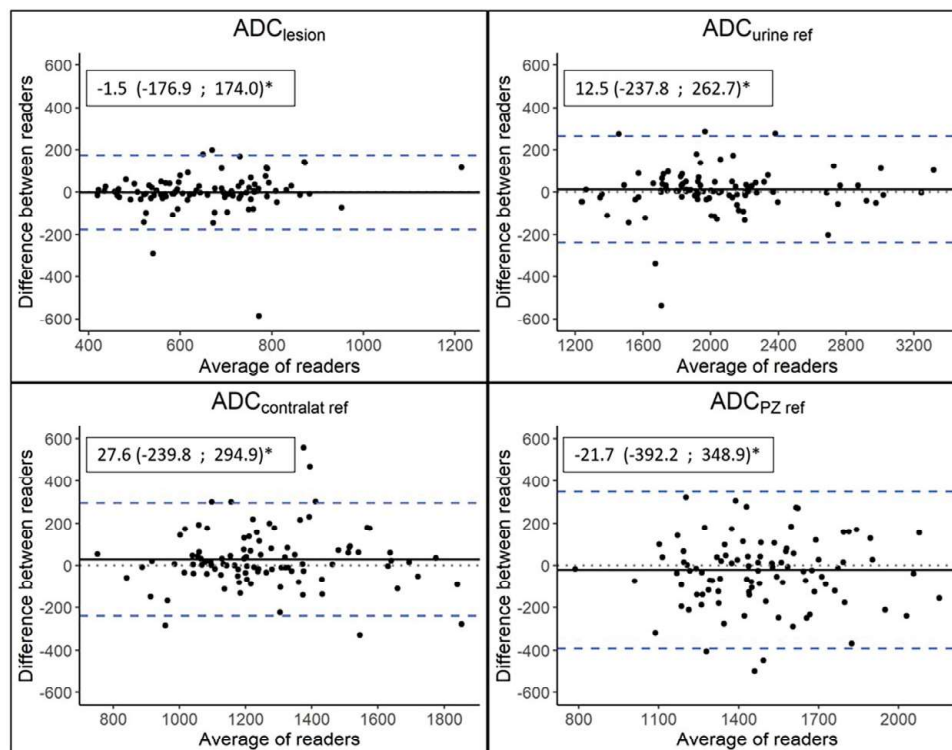


FIGURE 5
Bland-Altman plots. The dotted lines represent no difference between readers, the solid lines represent mean differences between readers, and the dashed blue lines represent limits of agreement, calculated as the mean difference \pm 1.96SD of the mean difference. *Mean differences between readers (95% limits of agreement).

Our study has several limitations. First, the study group was small. In addition, the quality of the MRI scans was generally lower than would have been acceptable today. Another limitation is that all included patients had cSPCa; therefore, we only obtained data from the more advanced and aggressive tumors. In contrast to previous articles on this topic, no patients with ISUP 1 tumors were subject to prostate resection. This is in line with current clinical treatment guidelines (35). Furthermore, we did not have information on the fraction of Gleason 4 in the ISUP 2 group (Gleason 3 + 4). A lower percentage of Gleason 4 could have put these patients in the group with non-significant cancers. Moreover, the results from pathology were extracted from the original pathology reports, which were produced in a clinical setting by different pathologists with different levels of experience. That is, no study-dedicated pathology examination was performed.

There is potential for improvement, which we will implement in a forthcoming study. Most important is to include the whole range of benign to the most aggressive tumors. This can be achieved by including core biopsies

performed using the MR – ultrasound fusion technique. Furthermore, with new digital pathology archives, high precision correlations can be made between the WM RALP specimen and corresponding MR slice. A dedicated reevaluation of a specific location in the WM specimen, including tumor subtype and tumor cell growth pattern, can be made.

Conclusions

In conclusion, our study did not find any correlation between the ADC value and ISUP grade in a multi-scanner setting. We found no benefit of using ADC ratios, so-called normalized ADC values, even with good agreement between the two experienced readers. This contradicts previous single-center studies published research in the field. Therefore, in a clinical situation with different MRI scanner types, measurements of ADC must be used with caution. It also highlights the importance of harmonizing the parameters of the MRI sequences across centers.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Lund University (Dnr 2014-886) Swedish ethical review authority (entry no. 2019-03674). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

JB, ET, DF-S, AB and SZ conceived of the presented idea. JB and ET performed the measurements. DFS retrieved clinical data from patient records. JB, ET, SZ processed the experimental data and performed the analyses. SZ and EB supervised the work. All authors contributed to the writing of the manuscript and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2023.1079040/full#supplementary-material>

References

1. Cancer IAFRo. *Global cancer observatory*. Lyon, France: International Agency for Research on Cancer (2020).
2. American College of Radiology Committee on PI-RADS® (Prostate). (2019). Available at: <https://www.acr.org/-/media/ACR/Files/RADS/PI-RADS/PI-RADS-V2-1.pdf> Accessed on October 1, 2022.
3. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA. The 2014 international society of urological pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma. *Am J Surg Pathol* (2016) 40(2):244–52. doi: 10.1097/PAS.0000000000000530
4. Kasivisvanathan V, Rannikko AS, Borghi M, Panebianco V, Mynderse LA, Vaarala MH, et al. MRI-Targeted or standard biopsy for prostate-cancer diagnosis. *N Engl J Med* (2018) 378(19):1767–77. doi: 10.1056/NEJMoa1801993
5. Barrett T, Priest AN, Lawrence EM, Goldman DA, Warren AY, Gnanapragasam VJ, et al. Ratio of tumor to normal prostate tissue apparent diffusion coefficient as a method for quantifying DWI of the prostate. *AJR Am J Roentgenol* (2015) 205(6):W585–93. doi: 10.2214/AJR.15.14338
6. Hambrook T, Somford DM, Huisman HJ, van Oort IM, Witjes JA, Hulsbergen-van de Kaa CA, et al. Relationship between apparent diffusion coefficients at 3.0-T MR imaging and Gleason grade in peripheral zone prostate cancer. *Radiology* (2011) 259(2):453–61. doi: 10.1148/radiol.11091409
7. Verma S, Rajesh A, Morales H, Lemen L, Bills G, Delworth M, et al. Assessment of aggressiveness of prostate cancer: Correlation of apparent diffusion coefficient with histologic grade after radical prostatectomy. *AJR Am J Roentgenol* (2011) 196(2):374–81. doi: 10.2214/AJR.10.4441
8. Boesen L, Chabanova E, Logager V, Balslev I, Thomsen HS. Apparent diffusion coefficient ratio correlates significantly with prostate cancer gleason score at final pathology. *J Magn Reson Imaging* (2015) 42(2):446–53. doi: 10.1002/jmri.24801
9. Hötker AM, Mazaheri Y, Aras Ö, Zheng J, Moskowitz CS, Gondo T, et al. Assessment of prostate cancer aggressiveness by use of the combination of quantitative DWI and dynamic contrast-enhanced MRI. *Am J Roentgenol* (2016) 206(4):756–63. doi: 10.2214/AJR.15.14912
10. Manetta R, Palumbo P, Giannarino C, Bruno F, Arrigoni F, Natella R, et al. Correlation between ADC values and Gleason score in evaluation of prostate cancer: Multicentre experience and review of the literature. *Gland Surg* (2019) 8(Suppl 3):S216–S22. doi: 10.21037/gs.2019.05.02
11. Hoeks CM, Hambrook T, Yakar D, Hulsbergen-van de Kaa CA, Feuth T, Witjes JA, et al. Transition zone prostate cancer: Detection and localization with 3-T multiparametric MR imaging. *Radiology* (2013) 266(1):207–17. doi: 10.1148/radiol.12120281
12. Costantini M, Belli P, Rinaldi P, Bufi E, Giardina G, Franceschini G, et al. Diffusion-weighted imaging in breast cancer: Relationship between apparent diffusion coefficient and tumour aggressiveness. *Clin Radiol* (2010) 65(12):1005–12. doi: 10.1016/j.crad.2010.07.008
13. Wu C-C, Jain R, Radmanesh A, Poisson LM, Guo W-Y, Zagzag D, et al. Predicting genotype and survival in glioma using standard clinical MR imaging apparent diffusion coefficient images: A pilot study from the cancer genome atlas. *Am J Neuroradiol* (2018) 39(10):1814–20. doi: 10.3174/ajnr.A5794
14. Surov A, Meyer HJ, Wienke A. Correlation between apparent diffusion coefficient (ADC) and cellularity is different in several tumors: A meta-analysis. *Oncotarget* (2017) 8(35):59492. doi: 10.18632/oncotarget.17752
15. Sun Y, Tong T, Cai S, Bi R, Xin C, Gu Y. Apparent diffusion coefficient (ADC) value: a potential imaging biomarker that reflects the biological features of rectal cancer. *PLoS One* (2014) 9(10):e109371. doi: 10.1371/journal.pone.0109371
16. Alessandrino F, Taghipour M, Hassanzadeh E, Ziaei A, Vangel M, Fedorov A, et al. Predictive role of PI-RADSv2 and ADC parameters in differentiating Gleason pattern 3 + 4 and 4 + 3 prostate cancer. *Abdom Radiol (NY)* (2019) 44(1):279–85. doi: 10.1007/s00261-018-1718-6
17. Rosenkrantz AB, Kopec M, Kong X, Melamed J, Dakwar G, Babb JS, et al. Prostate cancer vs. post-biopsy hemorrhage: Diagnosis with T2- and diffusion-weighted imaging. *J Magn Reson Imaging* (2010) 31(6):1387–94. doi: 10.1002/jmri.22172
18. Tamada T, Prabhu V, Li J, Babb JS, Taneja SS, Rosenkrantz AB. Assessment of prostate cancer aggressiveness using apparent diffusion coefficient values: Impact of patient race and age. *Abdom Radiol (NY)* (2017) 42(6):1744–51. doi: 10.1007/s00261-017-1058-y
19. Abreu-Gomez J, Walker D, Alotaibi T, McInnes MDF, Flood TA, Schieda N. Effect of observation size and apparent diffusion coefficient (ADC) value in PI-RADS v2.1 assessment category 4 and 5 observations compared to adverse pathological outcomes. *Eur Radiol* (2020) 30(8):4251–61. doi: 10.1007/s00330-020-06725-9
20. Woo S, Kim SY, Cho JY, Kim SH. Preoperative evaluation of prostate cancer aggressiveness: Using ADC and ADC ratio in determining Gleason score. *AJR Am J Roentgenol* (2016) 207(1):114–20. doi: 10.2214/AJR.15.15894
21. Bajgirani AM, Mirak SA, Sung K, Sisk AE, Reiter RE, Raman SS. Apparent diffusion coefficient (ADC) ratio versus conventional ADC for detecting clinically significant prostate cancer with 3-T MRI. *AJR Am J Roentgenol* (2019) 213(3):W134–W42. doi: 10.2214/AJR.19.21365
22. Ragheb SR, Bassiouny RH. Can mean ADC value and ADC ratio of benign prostate tissue to prostate cancer assist in the prediction of clinically significant prostate










- cancer within the PI-RADSv2 scoring system? *Egypt J Radiol Nucl Med* (2020) 51(1). doi: 10.1186/s43055-020-00347-3
23. Falaschi Z, Valenti M, Lanzo G, Attanasio S, Valentini E, Garcia Navarro LJ, et al. Accuracy of ADC ratio in discriminating true and false positives in multiparametric prostatic MRI. *Eur J Radiol* (2020) 128:109024. doi: 10.1016/j.ejrad.2020.109024
 24. Pepe P, D'Urso D, Garufi A, Priolo G, Pennisi M, Russo G, et al. Multiparametric MRI apparent diffusion coefficient (ADC) accuracy in diagnosing clinically significant prostate cancer. *vivo* (2017) 31(3):415–8. doi: 10.21873/invivo.11075
 25. Lebovici A, Sfrangeu SA, Feier D, Caraiani C, Lucan C, Suci M, et al. Evaluation of the normal-to-diseased apparent diffusion coefficient ratio as an indicator of prostate cancer aggressiveness. *BMC Med imaging* (2014) 14(1):1–7. doi: 10.1186/1471-2342-14-15
 26. De Cobelli F, Ravelli S, Esposito A, Giganti F, Gallina A, Montorsi F, et al. Apparent diffusion coefficient value and ratio as noninvasive potential biomarkers to predict prostate cancer grading: comparison with prostate biopsy and radical prostatectomy specimen. *AJR Am J Roentgenol* (2015) 204(3):550–7. doi: 10.2214/AJR.14.13146
 27. Litjens GJ, Hambrock T, Hulsbergen-van de Kaa C, Barentsz JO, Huisman HJ. Interpatient variation in normal peripheral zone apparent diffusion coefficient: Effect on the prediction of prostate cancer aggressiveness. *Radiology* (2012) 265(1):260–6. doi: 10.1148/radiol.12112374
 28. Itatani R, Namimoto T, Yoshimura A, Katahira K, Noda S, Toyonari N, et al. Clinical utility of the normalized apparent diffusion coefficient for preoperative evaluation of the aggressiveness of prostate cancer. *Japan J Radiol* (2014) 32(12):685–91. doi: 10.1007/s11604-014-0367-0
 29. Zhang J, Tian W-Z, Hu C-H, Niu T-L, Wang X-L, Chen X-Y. Age-related changes of normal prostate: Evaluation by MR diffusion tensor imaging. *Int J Clin Exp Med* (2015) 8(7):11220.
 30. Le Bihan D. What can we see with IVIM MRI? *Neuroimage* (2019) 187:56–67. doi: 10.1016/j.neuroimage.2017.12.062
 31. Gudbjartsson H, Patz S. The rician distribution of noisy MRI data. *Magn Reson Med* (1995) 34(6):910–4. doi: 10.1002/mrm.1910340618
 32. Surov A, Meyer HJ, Wienke A. Correlations between apparent diffusion coefficient and Gleason score in prostate cancer: A systematic review. *Eur Urol Oncol* (2020) 3(4):489–97. doi: 10.1016/j.euo.2018.12.006
 33. Shukla-Dave A, Obuchowski NA, Chenevert TL, Jambawalikar S, Schwartz LH, Malyarenko D, et al. Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. *J Magn Reson Imaging* (2019) 49(7):e101–e21. doi: 10.1002/jmri.26518
 34. Park KJ, Kim MH, Kim JK. Extraprostatic tumor extension: Comparison of preoperative multiparametric MRI criteria and histopathologic correlation after radical prostatectomy. *Radiology* (2020) 296(1):87–95. doi: 10.1148/radiol.2020192133
 35. Mottet N, Van den Bergh R, Briers E, Cornford P, De Santis M, Fanti S, et al. EAU-EANO-ESTRO-ESUR-SIOG guidelines on prostate cancer. *Eur Assoc Urol* (2020) 1–182.

Paper IV



Original Article

A pilot study of an organised population-based testing programme for prostate cancer

Max Alterbeck^{1,3} , Erik Thimansson^{2,4} , Johan Bengtsson^{2,4} , Erik Baubeta^{2,4} , Sophia Zackrisson^{2,4} , Anetta Bolejko^{2,4} , Kevin Sandeman⁵ , Sigrid Carlsson^{6,7,8} , Thomas Jiborn^{1,3} and Anders Bjartell^{1,3} 

¹Department of Urology, ²Department of Medical Imaging and Physiology, Skåne University Hospital, ³Division of Urological Cancers, Department of Translational Medicine, ⁴Division of Diagnostic Radiology, Department of Translational Medicine, Lund University, ⁵Department of Clinical Pathology and Molecular Diagnostics, Medical Services, Malmö, ⁶Department of Urology, Institute of Clinical Sciences, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden, ⁷Department of Surgery (Urology Service), and ⁸Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Objective

To determine the feasibility of a digitally automated population-based programme for organised prostate cancer testing (OPT) in Southern Sweden.

Patients and Methods

A pilot project for a regional OPT was conducted between September 2020 and February 2021, inviting 999 randomly selected men aged 50, 56, or 62 years. Risk stratification was based on prostate-specific antigen (PSA) level, PSA density (PSAD), and bi-parametric prostate magnetic resonance imaging (MRI). Men with a PSA level of 3–99 ng/mL had an MRI, and men with elevated PSA level (≥ 3 ng/mL) had a urological check-up, including a digital rectal examination and transrectal ultrasonography (TRUS). Indications for targeted and/or systematic transrectal prostate biopsies were suspicious lesions on MRI (Prostate Imaging-Reporting and Data System [PI-RADS] 4–5) and/or PSAD > 0.15 ng/mL/mL. Additional indications for prostate biopsies were palpable tumours, PSA ratio < 0.1 , or cancer suspicion on TRUS. Patient selection, mail correspondence, data collection, and algorithm processing were performed by an automated digital management system. Feasibility is reported descriptively.

Results

A total of 418 men had a PSA test (42%), with increasing participation rates by age (50 years, 38%; 56 years, 44%; and 62 years, 45%). Among these, 35 men (8%) had elevated PSA levels (≥ 3 ng/mL: one of 139, aged 50 years; 10/143, aged 56 years; and 24/146, aged 62 years). On MRI, 16 men (48%) had a negative scan (PI-RADS < 3), seven men (21%) had PI-RADS 3, nine men (27%) had PI-RADS 4, and one man (3%) had PI-RADS 5. All men with PI-RADS 4 or 5 underwent prostate biopsies, as well as two men with PI-RADS 3 due to PSAD > 0.15 ng/mL/mL or a suspicious finding on TRUS. Prostate cancer was diagnosed in 10 men. Six men underwent active treatment, whereas four men were assigned to active surveillance.

Conclusion

Our OPT model is feasible from an operational point of view, but due to the limited scale of this study no conclusions can be made regarding the efficacy of the diagnostic model or outcome.

Keywords

prostate cancer, screening, prostate-specific antigen, magnetic resonance imaging, algorithm

Introduction

Prostate cancer (PCa) continues to be the leading cause of cancer-related death for males in Sweden, as it has been for decades [1]. In the last two decades, improvements in diagnostics and curative and palliative treatments for these

patients have improved outcomes regarding PCa-specific morbidity and mortality. However, the incidence and prevalence of PCa are increasing, mainly due to improved diagnostics, a growing elderly population and men living longer with the disease. PCa is a heterogeneous disease with a spectrum from highly aggressive forms that metastasise early

to indolent forms that do not necessarily have the potential to metastasise and rarely lead to any symptoms or death. A major challenge is timely diagnoses, to find potentially lethal cancers at a stage when they may be successfully treated while avoiding the detection of cancers that are unlikely to become malignant, thereby preventing anxiety and complications from unnecessary treatments. The European Randomised Screening Study for Prostate Cancer (ERSPC) has demonstrated that screening, based on repeated PSA blood tests, reduces PCa mortality although at the cost of overdiagnoses and overtreatment [2]. Recently published STHLM3-MRI (ClinicalTrials.gov Identifier: NCT03377881) and Göteborg-2 (International Standard Randomised Controlled Trial Number ISRCTN94604465) trials in Sweden suggest that additional blood tests and MRI in the diagnostic evaluation may be included to avoid overdiagnoses and overtreatment [3]. The European Association of Urology (EAU) recommends the implementation of organised programmes for risk-stratified early detection of PCa that include MRI [4]. In June 2018, the public healthcare services committee in Southern Sweden (Region Skåne [RS]) was commissioned to implement a population-based organised PCa testing (OPT) programme. The focus of this OPT was to improve the availability, quality, and equality of PCa testing within a public healthcare setting. In contrast to current opportunistic PSA testing in primary care, all men in the region within the specific age groups will receive an OPT invitation, available in multiple languages, which will improve accessibility for underserved groups.

The context of regional OPT development, the implementation process, and structure of the digitalised population-based OPT programme in RS and Region Västra Götaland (VGR) have recently been described by Alterbeck *et al.* [5]. Prior to the implementation of the OPT programme in RS, a pilot project was conducted to assess the functionality of an automated digital OPT in terms of invitations, participation rates, and follow-up.

This pilot study was designed to compare participation rates, PSA outcomes, and PCa incidence in men aged 50, 56, and 62 years. Our data provide important insights that may inform resource utilisation and allocation for a large-scale programme targeting early detection of PCa. In the present study, we report the outcomes of our pilot project for OPT in Southern Sweden.

Patients and Methods

This pilot study was conducted between September 2020 and February 2021 in RS, the Southern County of Sweden, after approval by the Swedish Ethical Review Authority (2020-03923 and 2021-06647-02). Patient selection, mail correspondence, data collection, and algorithm processing were performed by the automated digital management system and supported by the OPT Head Office, as described

previously by Alterbeck *et al.* [5]. From the Swedish population registry, a total of 999 men aged 50 ($n = 367$), 56 ($n = 327$) or 62 years ($n = 305$) were randomly selected from 33 municipalities in RS to generate a representative cohort. An invitation letter, detailed information about the potential advantages and disadvantages of participating in PCa testing, a personal referral for PSA testing, and a research consent form were sent to each potential participant by mail. Blood samples for PSA testing were collected at primary care units or hospitals and sent for analysis at regional laboratories. PSA levels were automatically recorded in the administrative system and in each patient's medical record [5].

Management Algorithm, Including PSA and MRI

An algorithm for invitations and further management is illustrated in Fig. 1. A PSA level of < 3 ng/mL was considered negative and resulted in an automated response letter sent by mail; these men were reassigned to the OPT watchlist for new invitations after 2 years (if the PSA level was 1–2.9 ng/mL) or 6 years (if the PSA level was < 1 ng/mL). Men with a PSA level > 100 ng/mL were immediately referred for a urological assessment, with no MRI performed. A PSA level of 3–99 ng/mL resulted in a referral for prostate MRI at the nearest participating radiology department. An OPT workgroup of experienced radiologists agreed on MRI protocol requirements and diagnostic evaluation, in accordance with the Prostate Imaging-Reporting and Data System (PI-RADS) 2.1 document [6]. In addition, two expert radiologists performed a central review of all MRI examinations, for quality assurance. The standard procedure used was bi-parametric MRI (including T2-weighted and diffusion-weighted imaging sequences), and the radiologists' reports included prostate volume calculations for PSA density (PSAD), as well as focal lesion characteristics (PI-RADS 1–5) and locations (on a sector-based biopsy map).

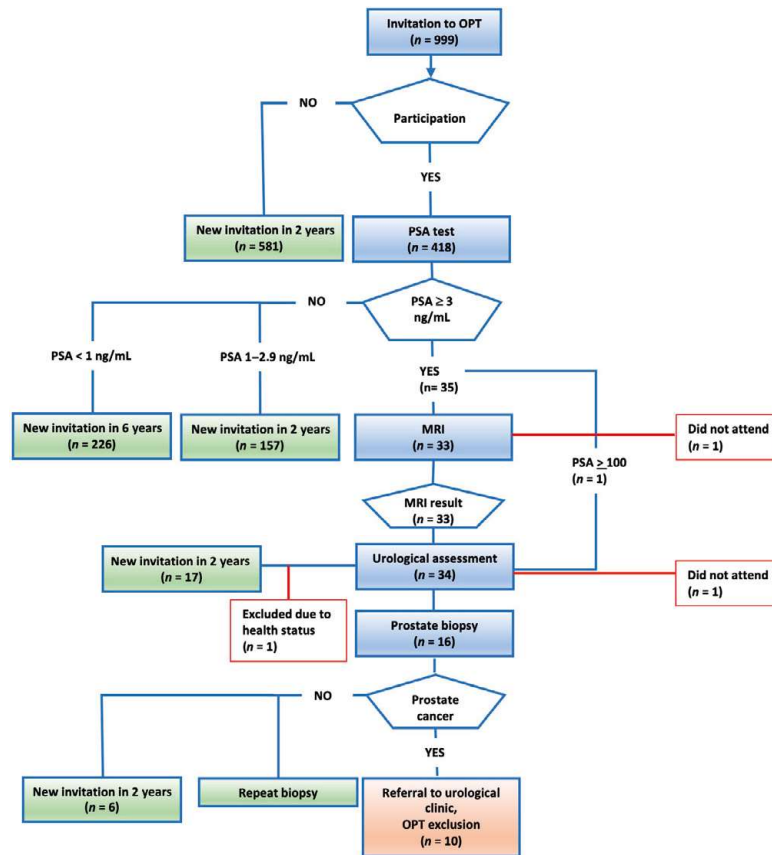
Urological Examination and Prostate Biopsy

All men with PSA levels of ≥ 3 ng/mL were referred to the nearest participating urology department (Malmö, Helsingborg, Ystad, Kristianstad, Landskrona, or Trelleborg). The urological assessment included a DRE and TRUS, as well as systematic, targeted, or combined prostate biopsies, as indicated by the algorithm (Fig. 2) [5]. Participants also completed a brief questionnaire at the urology department before their clinical visit, to provide information about their general health, family history of PCa, and use of anticoagulant medication.

Pathology

All biopsies were sent to regional pathology departments in Malmö, Helsingborg, or Kristianstad for histological

Fig. 1 Invitations, participation, and outcomes for the OPT pilot project. Men aged 50, 56, and 62 years were randomly invited to participate in OPT with an initial PSA test. Men with PSA levels of 3–99 ng/mL were automatically referred for bi-parametric MRI and subsequently for urological assessment with DRE and TRUS. Targeted and/or systematic TRUS-guided biopsies were taken if indicated by the OPT pilot algorithm (Fig. 2). Men with PSA levels of < 3 ng/mL were scheduled for another invitation after 2 years (PSA level = 1–2.9 ng/mL) or 6 years (PSA level < 1 ng/mL).



evaluation, based on the International Society of Urological Pathology (ISUP) 2014 consensus and the ISUP Gleason Grade Group (GG) classification [7].

Management of PCa/Further Investigation

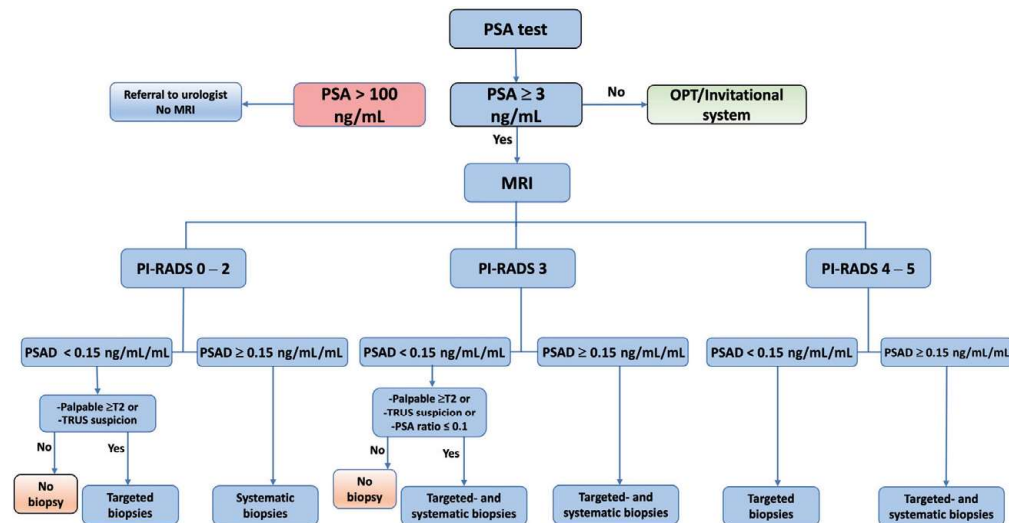
Patients with PCa-positive prostate biopsies were transferred to urology departments in RS for further evaluation and treatment, and by default excluded from the OPT programme. Biopsy-negative men were informed of their results by mail and returned to the OPT watchlist for a new invitation after 2 years with the following exceptions: (i) men

with PI-RADS > 4 underwent targeted biopsies with the TRUS/MRI-fusion technique; (ii) men with PI-RADS < 3 and PSAD > 0.3 ng/mL/mL had another PSA test after 3 months and if their PSA level was unchanged or increased, they had systematic and, if possible (i.e., PI-RADS 3), targeted biopsies with the TRUS/MRI-fusion technique (Fig. 3).

Statistical Analysis

Data were retrieved from the administrative system and medical records, including the municipal registry, number of

Fig. 2 The OPT pilot algorithm for men with elevated PSA levels. Men with elevated PSA levels (≥ 3 ng/mL, but < 100 ng/mL) underwent bi-parametric prostate MRI scans, which were assessed using PI-RADS. MRI was followed by urological assessments using DRE and TRUS. Targeted (three to four per lesion) and/or systematic (10–12) transrectal biopsies were taken when indicated by the PI-RADS assessments, PSAD (ng/mL/mL), or suspicious findings on TRUS or DRE. Men with PSA levels of ≥ 100 ng/mL were immediately referred for urological assessments and prostate biopsies without prior MRI.



invitations, participation rate, lead times within the OPT pathway, PSA levels, MRI results, PSAD values, cancer detection rates on prostate biopsy, and treatment choice. Descriptive statistics were reported as means, medians, interquartile ranges (IQRs), and percentages with binomial exact 95% CIs.

Results

Participation Rate

Among the 999 men who were invited, 418 (42%) participated and had a PSA test. In men aged 50, 56, and 62 years, the participation rates were 139/367 (38%), 143/327 (44%), and 136/305 (45%), respectively (Table 1).

Automated Digital Management System

All 999 invitations were managed by the digital management system and OPT head office, as were data retrieval (PSA, MRI, and biopsy results) from the 418 participants, adherence to the algorithm, and management of all response letters and future invitations.

Distribution of PSA Levels

The PSA levels among the 418 participants are shown in Table 1. In total, 35 participants (8%, 95% CI 6%–11%) had

PSA levels of ≥ 3 ng/mL with the following distribution among the age groups: one of 139 men aged 50 years (0.7%, 95% CI 0.02%–4%); 10/143 men aged 56 years (7%, 95% CI 3%–12%); and 24/136 men aged 65 years (18%, 95% CI 12%–25%). Men with elevated PSA levels (≥ 3 ng/mL) had a median (IQR) PSA level of 4 (3.2–5.8) ng/mL, and men who were diagnosed with PCa had a median (IQR) PSA level of 5.6 (3.43–36.5) ng/mL. The number of men with low (1–2.9 ng/mL) and very low (< 1 ng/mL) PSA levels decreased in the older age groups.

Magnetic resonance imaging

In total, 33 men with elevated PSA levels had MRI. Among these, 16 men (48%, 95% CI 31%–66%) had a negative MRI result (PI-RADS ≤ 2) and 17 men (52%, 95% CI 34%–69%) had a positive MRI result. Among the men with positive MRI results, seven examinations (21%, 95% CI 9%–39%) were classified as PI-RADS 3, nine (27%, 95% CI 13%–46%) as PI-RADS 4, and one as PI-RADS 5. Two men did not have MRI scans despite having elevated PSA levels: one man did not respond to his invitation for an MRI or for a urological examination despite multiple reminders, and one man had a PSA level of > 100 ng/mL and was immediately referred for urological assessment and a prostate biopsy. The median

Fig. 3 A 56-year-old man, PSA level = 3.1 ng/mL, PSAD = 0.10 ng/mL/mL. The MRI shows a PI-RADS 4 lesion in the dorsal portion of the mid-gland peripheral zone (PZ). Targeted biopsies (four) with cancer in all four biopsies, Gleason 3 + 4 = 7. RARP with the index tumour corresponding to a PI-RADS 4 lesion. **(A)** MRI diffusion-weighted imaging (DWI) b1500, white arrow indicates the lesion, grey arrow indicates artefact from rectal gas. **(B)** MRI apparent diffusion coefficient (ADC), arrow indicates the lesion. **(C)** MRI T2-weighted, arrow indicates the lesion. **(D)** Swedish nationwide web-based register platform for cancer patient data (INCA) database information from radiologist (lesion location is shown in yellow) and urologist (number and locations of targeted biopsies). **(E)** Digitalised pathology image of targeted biopsy shows 7-mm tumour, Gleason 3 + 4. **(F)** Digitalised image of a RP specimen. Green demarcation shows the index lesion (22 × 6 mm), dorsal portion of the mid-gland PZ. The second tumour, Gleason 3 + 3 (6 × 3 mm) anterior right side of the transition zone (TZ), was not detected by MRI.

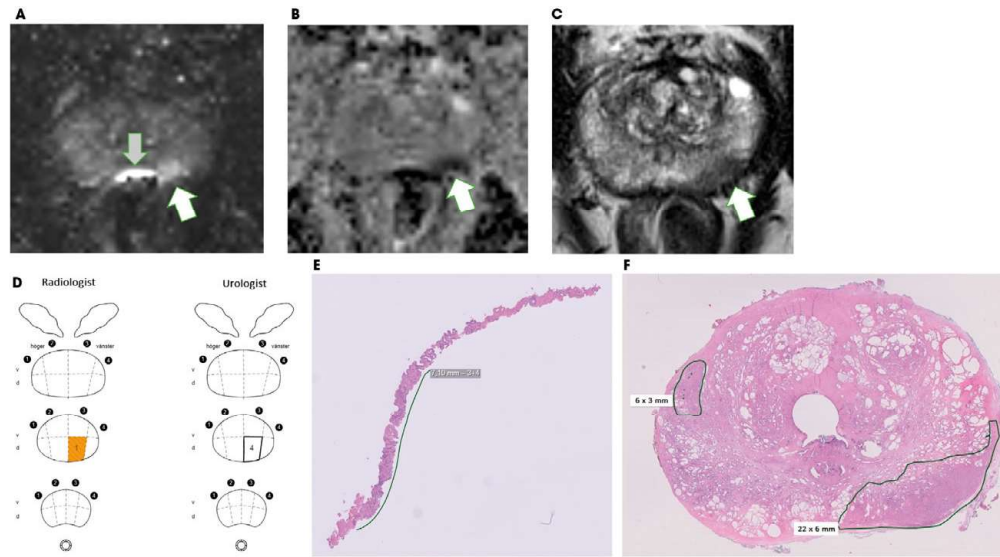


Table 1 The participation rates, PSA level results in ng/mL (%) and number of PCa diagnoses for men in the different age groups, as well as the number of men who return to OPT and are scheduled for another invitation after 2 years (PSA level = 1–2.9 ng/mL) or 6 years (PSA level < 1 ng/mL).

Variable	Age group			All men
	50 years	56 years	62 years	
Participation rate, n/N (%; 95% CI)	139/367 (38, 33–43)	143/327 (44, 38–49)	136/305 (45, 39–50)	418/999 (42, 39–45)
PSA < 1 ng/mL, n (%; 95% CI)	93 (67, 58–75)	77 (54, 45–62)	56 (41, 33–50)	226 (54, 49–59)
PSA = 1–2.9 ng/mL, n (%; 95% CI)	45 (32, 25–41)	56 (39, 31–48)	56 (41, 33–50)	157 (38, 33–42)
PSA ≥ 3 ng/mL, n (%; 95% CI)	1 (0.7, 0.02–4)	10 (7, 3–12)	24 (18, 12–25)	35 (8, 6–11)
PCa, n	0	1	9	10
Return to OPT after				
2 years, n (%; 95% CI)	45 (32, 25–41)	56 (39, 31–48)	56 (41, 33–50)	157 (38, 33–42)
6 years, n (%; 95% CI)	93 (67, 58–75)	77 (54, 45–62)	56 (41, 33–50)	226 (54, 49–59)
Total, n (%; 95% CI)	138 (99, 96–100)	133 (93, 88–97)	112 (82, 75–88)	383 (92, 89–94)

(IQR) time from the PSA test to the MRI report was 24 (13.5–29) days, and the median (IQR) time from the MRI report to the urological visit was 16 (12–20) days. The median (IQR) time between an invitation and a visit to a urological centre was 65 (52–76) days.

Prostate Biopsy Results and Treatments

A total of 16 men had prostate biopsies. In total, 13 of the 33 men who had MRI scans were biopsied based on PSAD values of > 0.15 ng/mL/mL and/or PI-RADS > 3. Two men

with PSAD values of < 0.15 ng/mL/mL and PI-RADS 3 lesions (i.e., no indications for biopsies, based on the OPT algorithm) had targeted biopsies due to suspicious findings on TRUS, and one man had biopsies without a prior MRI scan due to a PSA level of > 100 ng/mL.

Systematic prostate biopsies were taken in four participants, targeted biopsies in seven, and combined biopsies in five. The mean number of biopsies retrieved per participant was 10.8 for systematic biopsies, 4.5 for targeted biopsies, and 12.6 for combined biopsies.

Significant cancer, defined as ISUP GG ≥ 2 , was detected in seven of 10 men with PCa. The overall detection rates of PCa were 1% (10/999) and 2.4% (10/418) among men who were invited to participate and who had a PSA test, respectively. Clinically significant PCa was found in seven of 418 participants (1.7%), whereas three participants (0.7%) were diagnosed with clinically insignificant PCa (ISUP GG < 2).

The distribution of cancers in each PI-RADS group is shown in Table 2. One man had a high-grade prostatic intraepithelial neoplasia in all four of the four targeted biopsies from a PI-RADS 3 lesion and a PSAD < 0.15 ng/mL/mL; he was reassigned to the OPT watchlist for a new invitation after 2 years. Men with no cancer in their biopsy specimens were also reassigned to receive new invitations after 2 years.

Men diagnosed with PCa were managed according to national guidelines, with active surveillance (AS; four men), robot-assisted radical prostatectomy (RARP; three), and external beam radiation therapy (two), as well as androgen-deprivation therapy (one) for a case of metastasised disease. OPT associated characteristics for PCa-positive cases are shown in Table 3. According to the final pathology reports, all three men who underwent RARP had significant tumours that corresponded to the MRI lesions, in agreement with the biopsy findings. In addition, two of the men who started AS underwent RARP within 2 years, and in both cases, pathology reports showed GG2 and GG3 tumours corresponding to initial PI-RADS 4 lesions.

Discussion

This pilot project is the first study of risk-adapted OPT for early detection of PCa conducted in a public healthcare setting within Sweden. The project logistics worked well, and the number of participants and men diagnosed with PCa were as expected. Our study results should help to estimate participation rates and allocate necessary resources before launching a continuous regional or national OPT programme.

Recently, we published details of the design of our OPT programme, which was developed in parallel with a similar programme in VGR [5]. The VGR programme is a 3-year pilot project and the results have yet to be reported.

The overall participation rate in this pilot study was 42%, with increasing participation rates among older age groups. Our participation rate was somewhat lower than in earlier PCa screening studies in Sweden, such as the Göteborg-1 trial (ISRCTN54449243; 76%; participant age 50–64 years) and the Göteborg-2 trial (46%; participant age 50–60 years) [8,9]. However, lower participation rates have been noted in recent trials, such as the STHLM3-MRI study (26%; for men aged 50–74 years) and the Finnish ProScreen (NCT03423303) Pilot study (41%; for 65-year-old men) [3,10]. Because our pilot project was initiated in the autumn of 2020, restrictions due to the COVID-2019 pandemic may have affected the participation rate.

Details about the invitation to the OPT programme, and informed decision-making procedure, has been reported in detail previously. Participants were informed about the pros and cons of screening, in multiple available languages, and about the voluntary nature of participation [5]. One of the foundations of OPT is the decision of well-informed men to participate. As such, the participation rate reflects the rate of informed decision making, where choosing to undergo or forgo screening are both similarly accepted options, based on a man's values and preferences. However, if a low participation rate is due to structural or informational inadequacies of the programme, then future investigations

Table 2 Distribution of men with elevated PSA level, PSAD, the number of men who underwent biopsies, and the cancer-specific outcomes for those in each PI-RADS group and for those who did not undergo MRI.

Variable	Total, <i>n</i>	PI-RADS < 3	PI-RADS 3	PI-RADS 4	PI-RADS 5	MRI not performed, <i>n</i>
PSA level ≥ 3 ng/mL, <i>n</i> (%; 95% CI)	35	16 (48, 31–66)	7 (21, 9–39)	9 (27, 13–46)	1 (3, 0.08–16)	2*
PSAD > 0.15 ng/mL/mL, <i>n</i> (%; 95% CI)	6	3 (19, 4–45)	0 (0)	2 (22, 3–60)	1 (100, 3–100)	1
Biopsies, <i>n</i> (%; 95% CI)	16	3 (19, 4–45)	2 (29, 4–71)	9 (100, 66–100)	1 (100, 3–100)	1
PCa, <i>n</i> (%; 95% CI)	10	1 (6, 0.2–30)	1 (14, 0.4–58)	6 (67, 30–93)	1 (100, 3–100)	1
PSAD > 0.15 ng/mL/mL in cancer, <i>n</i>	6	1	0	1	1	1
GG > 2 , <i>n</i>	7	1	1	3	1	1

*One participant did not attend for their MRI scan or urological check-up.

Table 3 The OPT characteristics of PCa-positive individuals including PSA levels, PSA ratios (free/total), PSAD values, and T-stage. The table also indicates whether positive biopsies were found in MRI-positive lesions (PI-RADS > 3), the number of positive cores/total cores, and the total length of cancer-positive cores/total length of all cores, as well as subsequent treatment decisions.

Age, years	PSA level, ng/mL	PSA ratio (free/total)	PSAD, ng/mL/mL	PI-RADS score	T-stage	Biopsies	GG	Positive biopsy in lesion	Positive/total cores, n	Cancer/total length, mm	Treatment
62	4	0.24	0.06	4	T1c	Tbx	1	Yes	3/12	4.7/140	AS
62	5.6	0.1	0.18	4	T2a	SBx + TBx	1	Yes	2/16	0.8/180	AS*
62	5	0.17	0.08	4	T1c	Tbx	1	Yes	1/7	5/135	AS
62	3.2	0.11	0.11	4	T2b	Tbx	2	Yes	2/12	16.7/127	AS*
62	3.5	0.09	0.11	3	T1c	SBx + TBx	2	Yes	3/11	14/210	RP
62	38	n/a	0.29	<3	T1c	SBx	2	—**	1/12	2/173	RP
62	5.8	0.12	0.14	4	T1c	Tbx	2	Yes	3/4	7/64	EBRT
56	3.1	0.12	0.1	4	T1c	Tbx	2	Yes	4/4	25/74	RP
62	36	n/a	0.2	5	T1c	Tbx	5	Yes	6/6	43/110	EBRT
62	266	n/a	5.7	—***	T2	Sbx	5	—***	9/9	155/178	Hormone treatment

EBRT, external beam radiation therapy; n/a, not available; SBx, systematic biopsies; TBx, targeted biopsies. *Both cases underwent RARP within 2 years. **No lesion to target due to PI-RADS < 3. ***The participant did not have an MRI scan due to PSA level of > 100 ng/mL.

must examine the motives of non-attenders to improve the programme.

A strength of our study is the design involving three age groups, which provides information about the resources needed for screening men of different ages in terms of participation rates, PSA levels, number of MRI examinations, and the need for prostate biopsies. For example, in the 50-year-old age group, the participation rate was only 38%, and <1% of these participants had PSA levels of ≥ 3 ng/mL. By contrast, in the 62-year-old age group, the participation rate was 44%, and 18% of these participants had elevated PSA levels. Among the 10 men diagnosed with PCa, nine were detected in the 62-year-old age group and one in the 56-year-old age group. There is a tendency for the proportion of men with PSA levels of ≥ 3 ng/mL to increase with age, necessitating further examinations and re-invitations. These differences are of the utmost importance when planning and implementing an OPT programme because they affect the resources needed for MRI, prostate biopsies, subsequent treatments, and re-invitations.

Furthermore, because the PSA level has a tendency to fluctuate and many men return to 'normal' PSA-levels on subsequent measurement [11], it is possible that a second reflex-PSA could reduce unnecessary MRIs and prostate biopsies for men with temporarily elevated PSA levels due to e.g., inflammation or infection. However, there are practical challenges with such a reflex test in an automated OPT programme, mainly due to patient compliance, resource constraints and lead times issues.

In this pilot study, we found that 8% of the participants (all age groups) had elevated PSA levels of ≥ 3 ng/mL, which is consistent with the results reported for the Göteborg-2 trial, in which 7% of participants (aged 50–

60 years) had elevated PSA levels [9]. In the ProScreen trial, 17% of the men in the study group (aged 64–65 years) had elevated PSA levels, which is similar to our cohort, in which 18% of 62-year-old men had elevated PSA levels [10]. Based on our risk-stratified algorithm, which included PSAD and MRI assessments, approximately half of the men with elevated PSA levels (19/35) could be spared prostate biopsies.

Importantly, our PI-RADS frequencies should be interpreted with caution because there were comparatively few MRI examinations in our study. However, compared to other MRI screening studies, we observed a lower percentage (48%; 16/33) of negative MRI results (PI-RADS < 3) than in the Göteborg-2 (65%; 487/755) and Stockholm 3 (62%; 521/846) trials [3,9]. A significant advantage of incorporating MRI into a screening algorithm is avoiding unnecessary biopsies and overdiagnoses of indolent cancers, which in turn relies on avoiding false-positive MRI results. The proportion of PI-RADS 4 lesions in our pilot study was greater (27%; nine of 33) than in the Göteborg-2 (20%; 150/755) or STHLM3-MRI (10%; 85/846) trials. Furthermore, we found only three of nine of our PI-RADS 4 lesion biopsies had significant cancers, compared with 76% (65/85) of the PI-RADS 4 lesion biopsies in the Stockholm 3 trial. Notably, three other screening studies (Göteborg-2, STHLM2, and ProScreen) implemented centralised MRI reading by experienced radiologists, whereas in our pilot project, MRI reading was decentralised and performed by many readers with varying levels of experience, as would be the case in large population-based programmes. Further studies are needed to determine how MRI reading may be optimised in future screening programmes. Our research group plan to study the value of centralised MRI reading by expert radiologists in OPT in a future study.

Our PCa detection rate was 1% (10/999) among all men invited to participate in the study and 2.4% (10/418) among those men who did participate and had a PSA test. Clinically significant PCa was found in seven of 418 participants (1.7%), whereas three men (0.7%) were diagnosed with clinically insignificant PCa (defined as GG1). Our findings may be compared with those of the Göteborg-2 study, which identified 0.6% insignificant and 0.9% significant cancers in the experimental group [9]. The higher incidence of clinically significant cancers in our cohort may be because we invited a group of 62-year-old men and found nine of the 10 cancers in this group, whereas the men in the Göteborg-2 trial were aged 50–60 years.

Three of the 10 cancers detected were insignificant (GG1), compared with two of seven in the initial report from the ProScreen trial and 38% (66/176) in the Göteborg-2 study. By contrast, 18% (41/233) of cancers detected in the STHLM3-MRI trial were insignificant [3,9,10]. However, relatively few cancers were detected in either the ProScreen trial or in our study.

Previous studies have shown the importance of re-invitations in a screening programme [12]. Our algorithm includes re-invitations after 2 or 6 years, based on the findings of previous screening studies [2,8]. Delaying detection of localised PCa may not always be harmful and offering AS for localised PCa instead of immediate treatment may be a safe option, as demonstrated by the Scandinavian Prostate Cancer Group Study Number 4 (SPCG4) [13] and Prostate Testing for Cancer and Treatment (ProtecT) [14] studies.

Even if our algorithm misses some small medium-risk PCa, new assessments, which occur by default after 2 or 6 years, will probably detect any cancers that progress. Therefore, continuous evaluation of OPT outcomes and the occurrence of interval cancers is necessary to adapt the OPT algorithm and improve the detection of medium- and high-risk tumours.

Previous screening trials, such as the ESRPC, have shown that PCa screening programmes can increase the risk of overdiagnoses and overtreatment [2]. The risk of overdiagnoses can be reduced if algorithms for the diagnostic procedure include MRI, PSAD measurements, targeted biopsies, and other biomarker assessments. To reduce the risk of overtreatment, AS programmes may also be implemented. AS programmes have been successfully implemented in Sweden, where ~90% of patients with low-risk disease are managed [15]. We detected three ISUP GG1 tumours, which were managed by AS; one GG2 tumour was also managed by AS. However, two of four patients assigned to AS underwent RARP within 2 years, due to tumour upgrades at subsequent check-ups.

We included two precautionary procedures in our pilot study that would not be included in a full-scale OPT programme. First, all men with PSA levels of ≥ 3 ng/mL who underwent

MRI also had a urological assessment, regardless of MRI findings. Second, the urologist could recommend biopsies based on the TRUS findings, DRE, or free/total PSA ratio of < 0.1 . These precautionary procedures were implemented in the pilot study as quality assurance steps to ensure the integrity of the automated system and MRI evaluations. In total, two of 35 men with elevated PSA levels underwent biopsies due to these precautionary procedures. In one case, we found a small GG7 tumour; in the other case, we found a high-grade prostatic intraepithelial neoplasia. These two patients would not have been referred for biopsies under the full-scale OPT programme. Instead, they would have been reassigned to the OPT watchlist and given a new invitation after 2 years. These cases illustrate the trade-off between a sensitive testing algorithm and an effective but economical testing programme.

Due to the limited sample size ($n = 999$) and the short follow-up times, we cannot draw any conclusions regarding the efficacy of the OPT programme in reducing morbidity and mortality. However, the pilot study does provide valuable insights into the functionality of an automated and digitalised programme, as well as information on the resources needed to implement a full-scale OPT programme.

Conclusions

Our study showed that the proposed model and setup for a community-based, automated, risk-adapted OPT programme is feasible from an operational point of view. Our results suggest that an increasing age for invited men may impact participation rates, PSA levels, and PCa detection rates, affecting the resources needed for a full-scale OPT programme. Due to the limited scope of this study, it is not possible to make any assertions about the efficacy of OPT regarding the diagnostic algorithm, PCa mortality, morbidity, or cost-effectiveness.

Acknowledgements

The authors are grateful to NS Anna-Karin Börjedahl at the OPT head office and to research administrator Ms. Anna Holst for administrative work.

Disclosure of Interests

Sigrid Carlsson has received honorarium and travel reimbursements from Ipsen, unrelated to the present manuscript. Anders Bjartell has received the following which is not related to this article: honoraria, consulting fees, fees for serving on a speakers bureau, and/or travel support from Astellas Pharma, AstraZeneca, Bayer, Ferring Pharmaceuticals, IPSEN, Janssen, Merck Sharp & Dohme Corp, Sandoz and stock options from LIDDS Pharma, Glactone Pharma, Noviga Research and WntResearch. The other authors have no conflicts of interest to declare.

Funding

This work was supported by grants from the following: The Swedish Cancer Society (#CAN 2018/522); The Cancer Foundation at Skåne University Hospital Malmö; governmental funding (Avtal om Läkarutbildning och Forskning [ALF]) through The Faculty of Medicine at Lund University; a National Institutes of Health/National Cancer Institute Cancer Center Support Grant (P30-CA008748) to Memorial Sloan Kettering Cancer Center; and donations from Gösta Jönssons Fond, Hillevi Fries Fond and Prostatacancerförbundet.

References

- 1 National Board of Health and Welfare. *Cause of death statistics*. Stockholm: National Board of Health and Welfare, 2022. Available at: https://sdb.socialstyrelsen.se/lf_dor/val.aspx
- 2 Schröder FH, Hugosson J, Roobol MJ et al. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* 2009; 360: 1320–8
- 3 Nordström T, Discacciati A, Bergman M et al. Prostate cancer screening using a combination of risk-prediction, MRI, and targeted prostate biopsies (STHLM3-MRI): a prospective, population-based, randomised, open-label, non-inferiority trial. *Lancet Oncol* 2021; 22: 1240–9
- 4 Van Poppel H, Hogenhout R, Albers P, van den Bergh RCN, Barentsz JO, Roobol MJ. Early detection of prostate cancer in 2020 and beyond: facts and recommendations for the European Union and the European Commission. *Eur Urol* 2021; 79: 327–9
- 5 Alterbeck M, Järbur E, Thimansson E et al. Designing and implementing a population-based organised prostate cancer testing programme. *Eur Urol Focus* 2022; 8: 1568–74
- 6 Turkbey B, Rosenkrantz AB, Haider MA 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: 340–51
- 7 van Leenders G, van der Kwast TH, Grignon DJ et al. The 2019 International Society of Urological Pathology (ISUP) consensus conference on grading of prostatic carcinoma. *Am J Surg Pathol* 2020; 44: e87–99
- 8 Hugosson J, Carlsson S, Aus G et al. Mortality results from the Göteborg randomised population-based prostate-cancer screening trial. *Lancet Oncol* 2010; 11: 725–32
- 9 Hugosson J, Månsson M, Wallström J et al. Prostate cancer screening with PSA and MRI followed by targeted biopsy only. *N Engl J Med* 2022; 387: 2126–37
- 10 Rannikko A, Leht M, Mirtti T et al. Population-based randomized trial of screening for clinically significant prostate cancer ProScreen: a pilot study. *BJU Int* 2022; 130: 193–9
- 11 Eastham JA, Riedel E, Scardino PT et al. Variation of serum prostate-specific antigen levels: an evaluation of year-to-year fluctuations. *JAMA* 2003; 289: 2695–700
- 12 Arnsrud Godtman R, Holmberg E, Lilja H, Stranne J, Hugosson J. Opportunistic testing versus organized prostate-specific antigen screening: outcome after 18 years in the Göteborg randomized population-based prostate cancer screening trial. *Eur Urol* 2015; 68: 354–60
- 13 Bill-Axelsson A, Holmberg L, Ruutu M et al. Radical prostatectomy versus watchful waiting in early prostate cancer. *N Engl J Med* 2011; 364: 1708–17
- 14 Hamdy FC, Donovan JL, Lane JA et al. Fifteen-year outcomes after monitoring, surgery, or radiotherapy for prostate cancer. *N Engl J Med* 2023; 388: 1547–58
- 15 Sweden Tnpacro. NPCR: The national prostate cancer register of Sweden. 2022. Available at: <https://statistik.incanet.se/npcr/>. Accessed March 2023

Correspondence: Anders Bjartell, Department of Urology, Skåne University Hospital, Malmö, Sweden.

e-mail: anders.bjartell@med.lu.se

Abbreviations: AS, active surveillance; ERSPC, European Randomised Screening Study for Prostate Cancer; GG, Gleason Grade Group; IQR, interquartile range; ISRCTN, International Standard Randomised Controlled Trial Number; ISUP, International Society of Urological Pathology; OPT, organised PCa testing; PCa, prostate cancer; PI-RADS, Prostate Imaging-Reporting And Data System; PSAD, PSA density; PSA ratio, free PSA/total PSA; (RA)RP, (robot-assisted) radical prostatectomy; RS, Region Skåne; VGR, Region Västra Götaland.