



Published in final edited form as:

*Scand J Urol.* 2019 October ; 53(5): 304–311. doi:10.1080/21681805.2019.1675757.

## Performance and Inter-observer Variability of Prostate MRI (PI-RADS version 2) Outside High-volume Centres.

Kimia Kohestani<sup>\*,1,2</sup>, Jonas Wallström<sup>\*,3,4</sup>, Niclas Dehlfors<sup>4</sup>, Ole Martin Sponga<sup>5</sup>, Marianne Månsson<sup>1</sup>, Andreas Josefsson<sup>1,2,6,7</sup>, Sigrid Carlsson<sup>1,8</sup>, Mikael Hellström<sup>3,4</sup>, Jonas Hugosson<sup>1,2</sup>

<sup>1</sup>Department of Urology, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg Sweden

<sup>2</sup>Region Västra Götaland, Sahlgrenska University Hospital, Department of Urology, Gothenburg Sweden

<sup>3</sup>Department of Radiology, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg Sweden

<sup>4</sup>Region Västra Götaland, Sahlgrenska University Hospital, Department of Radiology Gothenburg Sweden

<sup>5</sup>Department of Radiology, Carlanderska Hospital, Gothenburg, Sweden

<sup>6</sup>Department of Surgical and Perioperative Sciences, Urology & Andrology, Umeå University, Sweden.

<sup>7</sup>Wallenberg Center for Molecular medicin, Umeå University, Umeå Sweden

<sup>8</sup>Department of Surgery (Urology Service) and Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, USA

### Abstract

**Objective:** Despite the growing trend to embrace pre-biopsy MRI in the diagnostic pathway for prostate cancer (PC), its performance and inter-observer variability outside high-volume centres remains unknown. This study aims to evaluate sensitivity of and variability between readers of prostate MRI outside specialized units with radical prostatectomy (RP) specimen as the reference standard.

**Materials and methods:** Retrospective study comprising a consecutive cohort of all 97 men who underwent MRI and subsequent RP between Jan 2012-Dec 2014 at a private hospital in Sweden. Three readers, blinded to clinical data, reviewed all images (including 11 extra prostate MRI to reduce bias). A tumour was considered detected if the overall PI-RADS v2 score was 3–5 and there was an approximate match (same or neighbouring sector) of tumour sector according to a 24 sector system used for both MRI and whole mount sections.

Corresponding author: Kimia Kohestani, Address: Bruna Stråket 11B, 413 45 Gothenburg, Sweden, Kimia.kohestani@vgregion.se.

\* Authors contributed equally

Conflict of Interest

Sigrid V. Carlsson has received a lecture honorarium and travel support from Astellas Pharma (unrelated to current study).

**Results:** Detection rate for the index tumour ranged from 67–76% if PI-RADS 3–5 lesions were considered positive and 54–66 % if only PI-RADS score 4–5 tumours were included. Detection rate for aggressive tumours (GS 4+3) was higher; 83.1 % for PI-RADS 3–5 and 79.2 % for PI-RADS 4–5. The agreement between readers showed average  $\kappa$  values of 0.41 for PI-RADS score 3–5 and 0.51 for PI-RADS score 4–5.

**Conclusions:** Prostate MRI evidenced a moderate detection rate for clinically significant PC with a rather large variability between readers. Clinics outside specialized units must have knowledge of their performance of prostate MRI before considering omitting biopsies in men with negative MRI.

### Keywords

prostatic neoplasms; magnetic resonance imaging; radical prostatectomy; interobserver agreement; observer variation

---

## Introduction

Prostate cancer (PC) forms a major public health concern in Western society[1, 2]. The standard method to diagnose PC has been transrectal ultrasound (TRUS)-guided systematic biopsies of the prostate. However, a major drawback with this approach is the risk of over-detection of clinically non-significant PC due to low specificity of elevated PSA and high prevalence of indolent tumours in the prostate[3, 4, 5, 6, 7]. The low specificity of PSA also implies that many men with elevated PSA-levels undergo unnecessary prostate biopsy, with its concomitant risks of bleeding, infection, and hospitalization. Hence, a better diagnostic approach is urgently needed. Therefore, pre-biopsy magnetic resonance imaging (MRI) of the prostate and performing (targeted) biopsies only in men with suspicious MRI is a rapidly growing trend in the diagnostic pathway[8].

Prostate MRI performed at high-volume centres and reported by experienced radiologists has shown promising results in reducing the number of unnecessary biopsies and reducing over-diagnosis[9]. The PRECISION trial, a multicentre study randomizing men to MRI with or without biopsies targeted to tumour-suspicious areas versus standard TRUS-guided systematic biopsies, showed that the MRI-arm was superior with more clinically significant PC detected, fewer men undergoing biopsy, and fewer clinically insignificant PC detected[10].

With increased use of prostate MRI outside specialized units it is relevant to question if the promising results from high-volume centres can be reproduced in a less experienced setting. In a routine clinical care study, substantial variation was reported in PI-RADS-score assignment and significant cancer yield among radiologists [11]. Diagnostic accuracy of pre-biopsy MRI in detecting clinically significant PC ranges from 44% to 87% and with a negative predictive value (NPV) ranging between 63% and 98%[12, 13].

Until recently, due to the large variability in NPV, most guidelines on PC did not recommend routine use of pre-biopsy MRI in biopsy-naïve patients[14]. Now, we find a shift, however. The updated 2019 EAU (European Association of Urology) Guidelines recommend

performing MRI in both biopsy-naïve patients and previously biopsied patients with clinical suspicion of PC [15]. There is an ongoing debate whether it is safe to omit biopsies in MRI-negative men.

To standardize interpretation and reporting of prostate MRI, the European Society of Urogenital Radiology (ESUR) and the American College of Radiology (ACR) developed the Prostate Imaging Reporting and Data System version 2 (PI-RADS v.2)[16]. PI-RADS v.2 has been shown to have low specificity[17, 18] and a moderate inter-reader reproducibility[18, 19, 20, 21], which are major concerns to consider before MRI can safely be used in routine.

The purpose of this study was to evaluate the performance and variability between readers of prostate MRI outside specialized units with prostatectomy (RP) specimen as the reference standard.

## Materials and methods

### Study design and participants

All men who consecutively underwent retropubic radical prostatectomy (RRP) or robotic-assisted radical prostatectomy (RARP) between January 1, 2012 and December 31, 2014 and who also underwent preoperative MRI of the prostate were identified (n=107). These men were treated at a private hospital in Sweden where pre-surgical MRI was incorporated into routine care early on. However, patients were selected for surgery based upon findings at TRUS-guided biopsies. MRI was performed in the majority of patients (73.2 %) before the diagnostic biopsies. MRI information was not used to select patients but for strategic planning of the surgery (e.g., degree of nerve-sparing). Of eligible patients, ten were excluded: 1 (neoadjuvant hormonal therapy); 1 (non-tumour finding at final pathology); 5 (MRI images unavailable for review; and 3 (poor quality of MRIs). To reduce confirmation bias, 11 additional MRIs judged as normal and from men with benign pathology at TRUS biopsies and/or TURP were added and randomly mixed to the list for review. Readers were informed of the incorporation of “normal” MRIs in the study but not the total number. These 11 extra cases were not included in the analysis, resulting in a final study population of 97 patients of which 84 were operated by RRP and 13 by RARP (Figure 1). The study was approved by The Regional Ethical Review Board in Gothenburg (registration number 515–14).

### MRI and review

MRI examinations were performed at 16 different hospitals (median time 1 month prior to RP). The majority of examinations (66 of 97) were performed using a 1.5-T scanner (GE Medical Systems Signa HDe) with a pelvic phased-array coil in the same hospital where the men underwent surgery. Pulse sequences included multiplanar fast spin echo (FSE) T2-weighted imaging, axial diffusion weighted imaging (DWI), and axial dynamic contrast enhanced (DCE) T1-weighted imaging with DCE imaging as the final sequence. DWI imaging included a high b value of 1500 s/mm<sup>2</sup>. Due to scanner restrictions, only two b-values were acquired (0 and 1500 s/mm<sup>2</sup>) and used to construct the apparent diffusion

coefficient (ADC)-map. DCE imaging was performed after 0.1 mmol/kg gadoterate meglumine (Dotarem, Gothia Medical), administered as an intravenous bolus injection via a power injector (Medtron, Gothia Medical).

Of 97 examinations, 12 were performed before the publication of PI-RADSv1 and all were performed without DCE. After PI-RADSv1 was published, the main site protocol was changed to include DCE. At the external sites, DCE was performed according to local preferences. One third of examinations (34 of 97) were bi-parametric and did not include DCE. MRI characteristics are reported in Supplementary table 1. Image interpretation was performed using SECTRA PACS and interpretation of DCE images was based on source images alone without post-processing.

A retrospective MRI review of all the images was performed and reported by three readers independently of each other. Readers were blinded to original MRI reports, clinical data, and pathology results. In addition, one reader (reader 1) performed a secondary analysis of pathology results after unblinding in cases assessed by all readers to be MRI-negative.

Each reader reported a maximum of three lesions per case using a structured reporting template based on the PI-RADS version 1 guidelines, except for lesion mapping. Lesions were mapped on a 24-sector template adopted in Sweden as the national standard for pathology and radiology reports (Figure 2). Following the review, reader 1 assigned each lesion to the relevant prostate zone before using a computer script converting the PI-RADSv1 to PI-RADSv2 overall scores. The MRI index lesion was defined as the lesion with the highest overall PI-RADSv2 score. In case several lesions with the same PI-RADS score were described, the largest one was considered the index lesion.

### **Radical Prostatectomy (RP) pathology, Prostate scheme, and correlation to MRI-lesions**

The RP specimens were processed and evaluated at the Department of Pathology, Unilabs Skövde (Skaraborgs Sjukhus, Skövde, Sweden). Size and Gleason Score (GS) of the three largest tumour foci on the RP specimens were recorded from the pathology reports and scanned whole mounts slides. This was done by one of the investigators (K.K), who also recorded the localization of each tumour according to the same 24-sector template as described above. If several tumours were found, the index tumour was defined as the tumour with the highest GS or the largest one if several tumours with the same GS were found. Two of the investigators (K.K and J.W) compared localization of the index tumour with MRI reports. The index tumours were considered detected if there was a lesion with a PI-RADS score 3–5 described in the same or neighbouring sector (approximate match).

### **Statistical analysis**

Descriptive statistics were applied to compare the demographics between groups of patients with MRI-negative tumours (defined as not assessed by any of the readers) and patients with MRI-positive tumours. Characteristics of the missed tumours are described. Detection rate of index tumour was calculated for each reader separately and for overall detection.

Inter-observer agreement between each pair of readers for lesions was evaluated using Cohen's  $\kappa$  coefficient and assessed according to Landish and Coch[22], considering slight

agreement as  $\kappa$  0.01–0.20, fair agreement  $\kappa$  0.21–0.40, moderate agreement  $\kappa$  0.41–0.60, substantial agreement,  $\kappa$  0.61–0.80, and almost perfect agreement  $\kappa$  0.81–0.99. The  $\kappa$  coefficient is an index of agreement corrected for chance and thus a high prevalence of a given observation can produce a counter-intuitive low  $\kappa$  coefficient even if the inter-observer agreement is almost perfect.

Analyses were performed using SPSS (version 24) and R Statistical Software (Version 3.3.1; Foundation for Statistical Computing, Vienna, Austria).

## Results

Patient- and tumour characteristics before and after RP are shown in Table 1.

There were no noteworthy differences between the groups of men with MRI-positive and MRI-negative index tumours (not identified by any of the readers) with respect to age, PSA level, prostate volume, or largest tumour diameter at pathology (Supplementary table 2). Median density (PSAD) was lower in the MRI-negative group, 0.11 (IQR 0,15; 0,35) compared to the MRI-positive group, PSAD 0.21 (IQR 0,08; 0,25).

Average index tumour detection rate for the readers was 73% (range 67% - 76%) if lesions with PI-RADS score 3–5 were considered and 61% (range 54%–66%) if PI-RADS score 4–5 were considered. Average detection rate for aggressive tumours (GS 4+3) was higher; 83% (range 77%–88%) for PI-RADS 3–5 and 79% (range 69%–85%) for PI-RADS 4–5 (Tables 2a *and* 2b).

There was fair to moderate agreement for pairwise combination of readers with lesions PI-RADS 3 with  $\kappa$  coefficients of 0.38 (95% CI 0.18–0.58), 0.31(95% CI 0.09–0.53), and 0.54 (95% CI 0.35–0.72), and moderate agreement for pairwise combinations of readers with lesions PI-RADS 4 with  $\kappa$  coefficients of 0.50 (95% CI 0.32–0.67), 0.49 (95% CI 0.31–0.68), and 0.54 (95% CI 0.37–0.71) (Table 3).

In 11 men, index tumour was missed by all readers: three men had GS 3+3 (including one man with 2 mm GS 3+3); six men had GS 3+4; one man had GS 4+3 and one man had GS 4+5. Complete characteristics of these index tumours that no reviewer reported are shown in Table 4a. Three of these men had a second or third significant tumour correctly identified by all readers including the man with GS 4+5.

Of the missed index tumours, five were located in the apex, two in the mid gland, and four in the base of the prostate. Nine were peripheral zone tumours.

After unblinding of pathology results, reader 1 could retrospectively identify four of 11 previously missed index tumours (1 centrally around urethra; 1 in transition zone; 2 in peripheral zone). Further, another 11 men had a MRI where only one of the readers reported the index tumour, Table 4b.

There were 23 false positive MRI lesions in the group of 97 men with proven PC. False positives were defined as PI-RADS 3–5 without any correlating tumour in the prostatectomy

specimen. Of these, 2 lesions were scored PI-RADS 5, 13 lesions PI-RADS 4, and 8 lesions PI-RADS 3. The number of false positive lesions per reader were 6 lesions (reader 1), 5 lesions (reader 2), and 12 lesions (reader 3), with a positive predictive value (PPV) of 92%, 92%, and 86%, respectively.

In the sample of 11 men with benign histology, reader 1 reported one suspicious lesion (PI-RADS 3–5), reader 2 reported two lesions, and reader 3 reported eight lesions.

## Discussion

The aim of this study was to evaluate performance and inter-observer variability between readers of prostate-MRI outside high-volume clinics, using histology from RP as reference standard. Detection rate for the index tumour ranged from 67–76% if PI-RADS 3–5 lesions were considered positive and 54–66 % if only PI-RADS score 4–5 tumours were included. Detection rate for aggressive tumours (GS 4+3) was higher; 83.1 % for PI-RADS 3–5 and 79.2 % for PI-RADS 4–5.

In a study by Greer et al., comparing detection of index lesions across five body radiologists from two different institutions with prostatectomy specimens as reference standard, the average sensitivity for detecting index lesions defined as GS7 or above was 91% [23]. This study included more high-grade PC (approximately 50% GS 4+3 or above), all imaging was performed at 3T at a single institution and used moderately to highly experienced body radiologists, which may account for the higher reported detection rate. In a meta-analysis including 526 patients from seven studies, the pooled sensitivity for MRI detection of clinically significant PC was lower, 74% [24].

Regarding the 11 men whose index tumour was not identified by any of the readers, three had GS 3+3 tumours (27%) and one of these tumours was only 2 mm in diameter and thus considered clinically insignificant. The remaining eight men all had GS 3+4 and seven of these had multifocal disease. Of those with multifocal disease, another significant tumour (GS 3+4) was correctly identified by all readers. Borofsky et al. reported a prospective detection rate of 84% for the two largest clinically significant lesions when correlating MRI reports to prostatectomy specimens [25]. This study included 162 lesions in 100 patients corresponding to 1 GS 3+3 tumour, 97 GS 3+4/4+3 tumours, and 64 GS 8 tumours. When re-reading the MRI, a majority of missed tumour lesions were multifocal and 17 out of the 26 missed lesions were GS 3+4 (65%). None were GS 6. Like our study, a lower PSAD was noted in the group of men with missed tumours. Multi-focality and low PSAD might be related to difficulties of visualizing tumours on MRI but this needs to be further studied.

There was a large variation in the number of false positives reported by the three readers. The reader with the highest number of false positives also had the best index tumour detection rate. At the same time, this reader reported over 70% PI-RADS 3–5 lesions in the added sample with negative MRI:s, indicating that this reader had a bias towards overcalling lesions.

We report fair to moderate pairwise inter-observer agreement for lesions stratified as either PI-RADS 3 or PI-RADS 4 with a  $\kappa$  of 0.41 and 0.51. In a study by Muller et al., a

moderate  $\kappa$  value of 0.46 for scoring of predefined suspicious lesions according to PI-RADS v.2 was reported [21]. However, in our study, suspicious lesions were not presented to the radiologists beforehand, adding another level of authenticity to resemble the daily workflow. In an article by Sonn et al., a substantial variation was found across radiologists with different reporting volumes in a specialized centre reporting according to PI-RADS v.2, concluding that internal validation of MRI is warranted before widespread adoption [11]. In our study, we found a moderate agreement level between readers despite a heterogeneous MRI population, varying radiologist experience, and a low-volume setting.

Interestingly, Rosenkrantz et al. found that the learning curve in prostate tumour detection among six second-year residents largely reflected self-directed learning with less effect of continual feedback and a plateau after a moderate number of cases, approximately 40 [26]. In our study, the two board-certified readers who had read > 200 cases performed somewhat better (average detection rate 76%) than the resident who had read approximately 50 cases (average detection rate 67%), supporting that only a moderate number of cases are required to reach this learning plateau. Whether very high experience by readers can further improve the performance could not be answered in our study and if advances in technology of MRI readings could further enhance the results remains to be evaluated.

Our study has several limitations. First, the study design was retrospective and readers were aware that most of the men had been subjected to prostatectomy. However to reduce this potential bias, 11 biopsy-negative MRIs were added to the reading list. We acknowledge that this does not represent a clinical pre-biopsy situation but the reason for adding negative MRI:s was to reduce reader bias, not to create a substitute for a pre-biopsy setting. The biopsy-negative MRIs were excluded from the final analysis as they do not represent prostatectomy-confirmed true negatives.

Second, MRI was performed at different institutions with different protocols. The MRI protocol at the main site only included DWI imaging at b0 and b1500 due to hardware restrictions in MRI scanner. When calculating ADC-maps PI-RADSv2.1 recommends using a maximum b-value of 1000 mm/s<sup>2</sup> to minimize the departure from basic monoexponential diffusion. We consider using b0 and b1500 for constructing the ADC-map a minor limitation. If ADC image data is used for visual inspection only and not for quantitative thresholding and analysis, tumour conspicuity will not decrease since ADC values will be lower on both tumour and normal peripheral zone. A bigger problem is low signal to noise when performing DWI at high b-values. A small number of examinations (3 in total) were excluded due to low DWI quality.

Third, a rather large proportion of examinations were performed without DCE (31%). About 1/3 of these MRI:s were performed early in the study before the main site adopted mpMRI according to ESUR recommendations. There is a controversy regarding the diagnostic accuracy of bi-parametric prostate MRI (without DCE) in two recent meta-analyses reporting a slight advantage for multi-parametric MRI and no advantage, respectively [27, 28]. However, the heterogeneity of the MRI population may be a strength since many hospitals in Scandinavia are not using DCE which brings this study closer to clinical routine.

For the moderately experienced readers in this study, including bi-parametric MRI may have negatively affected cancer detection rates.

Fourth, although the PI-RADSv2 guidelines were recently published at the time of the retrospective review, it was decided in the study group to continue scoring each pulse sequence individually in the fashion of PI-RADSv1. However, the readers were aware of the update and the pictorial examples in the PI-RADSv2 document were used for reference. Finally, we did not include any expert readers as reference but this may be a future study and a prospective study of low volume institutions with targeted biopsies as reference standard to evaluate if targeted biopsies outperform standard biopsies also in routine care.

In conclusion, in this study, prostate MRI evidenced a moderate detection rate for clinically significant PC. There is a rather large variability between readers which decreases if only PI-RADS 4 and 5 are considered. The variability outside specialized units makes it important for each unit to evaluate their own performance before they can safely use MRI in clinical routine to exclude men from undergoing prostate biopsies.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

This work was supported by research grants from the Swedish Cancer Society (Contract number 2017/620), The Swedish Research Council (no. 2016–01973), and from the Swedish state under the agreement between the Swedish government and the county councils, the ALF-agreement (ALFGBG-724401 and ALFGBG-774531). Part of Kimia Kohestani's work on this paper was supported by Anna-Lisa and Bror Björnsson's Foundation, Märta and Gustaf Ågren's Research Foundation, the Research Foundation at the Department of Urology at Sahlgrenska University Hospital, and the Royal and Hvitfeldtska Foundation. Sigrid V. Carlssons work on this paper was supported by research grants from the Sidney Kimmel Center for Prostate and Urologic Cancers, a Specialized Program of Research Excellence grant (P50-CA92629) from the National Cancer Institute to Dr. Howard Scher, a National Institutes of Health/National Cancer Institute Cancer Center Support Grant (P30-CA008748) to Memorial Sloan Kettering Cancer Center, a grant from the National Cancer Institute as part of the Cancer Intervention and Surveillance Modelling Network (U01CA199338–02), and the David H. Koch prostate cancer research fund. None of the sponsors had any part in the study design or access to the data. Nor had they any influence on or access to the analysis, the results, or the manuscript. We sincerely thank Ewa Löfkvist for her kind assistance in retrieving the pathology reports and Helén Ahlgren for her excellent help with database management. We also thank Karin Stinesen Kollberg for English language editing and review.

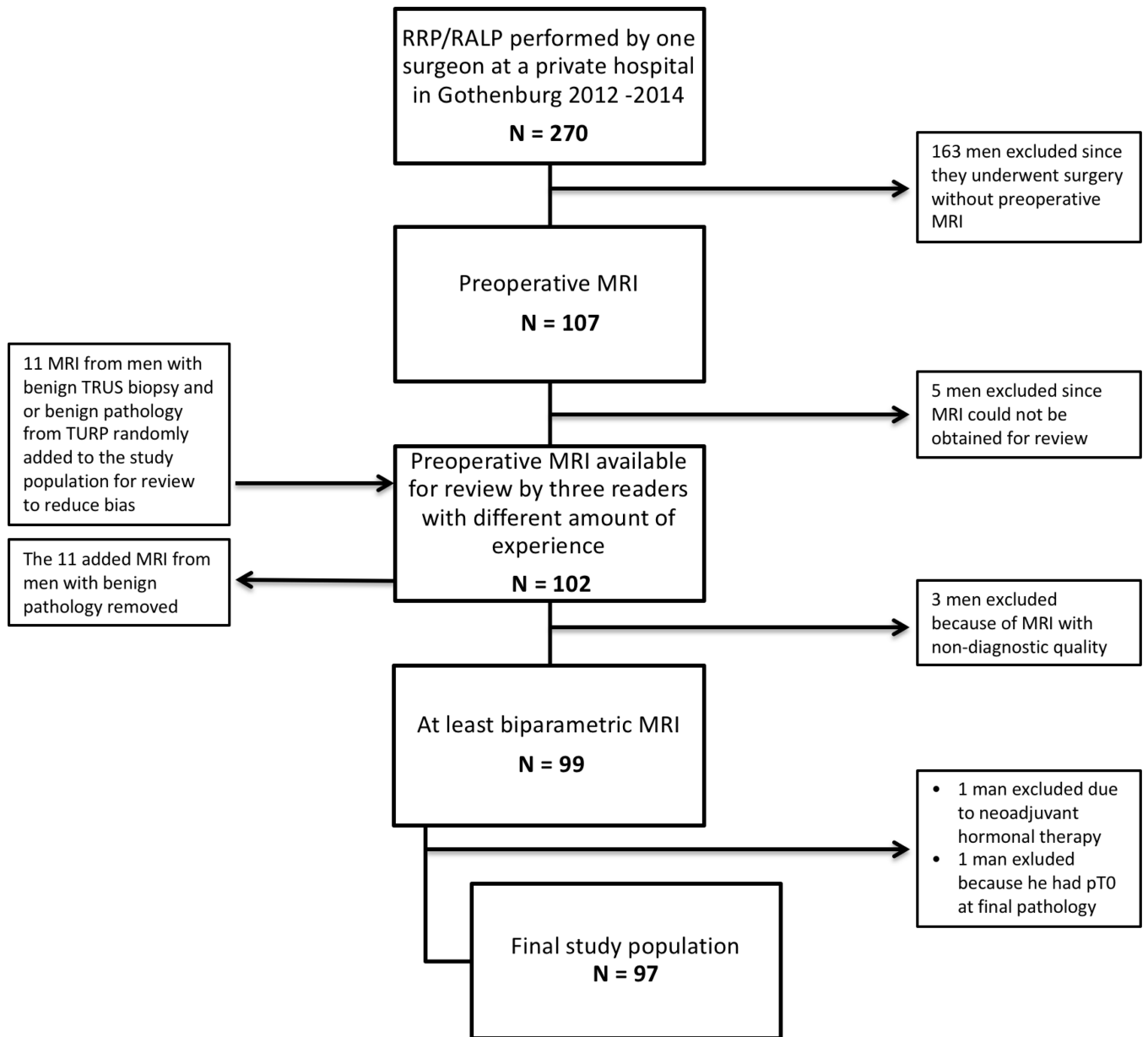
## References

1. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer*. 2013 4;49(6):1374–403. doi: 10.1016/j.ejca.2012.12.027. PubMed PMID: . [PubMed: 23485231]
2. The National Board of Health and Welfare Official statistics of Sweden. Statistics on cancer incidence 2016 in Sweden 2017, December 31. Available from: <https://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/20787/2017-12-31.pdf>
3. Draisma G, Boer R, Otto SJ, et al. Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. *J Natl Cancer Inst*. 2003 6 18;95(12):868–78. PubMed PMID: ; eng. [PubMed: 12813170]
4. Sanchez-Chapado M, Olmedilla G, Cabeza M, et al. Prevalence of prostate cancer and prostatic intraepithelial neoplasia in Caucasian Mediterranean males: an autopsy study. *Prostate*. 2003 2 15;54(3):238–47. doi: 10.1002/pros.10177. PubMed PMID: . [PubMed: 12518329]

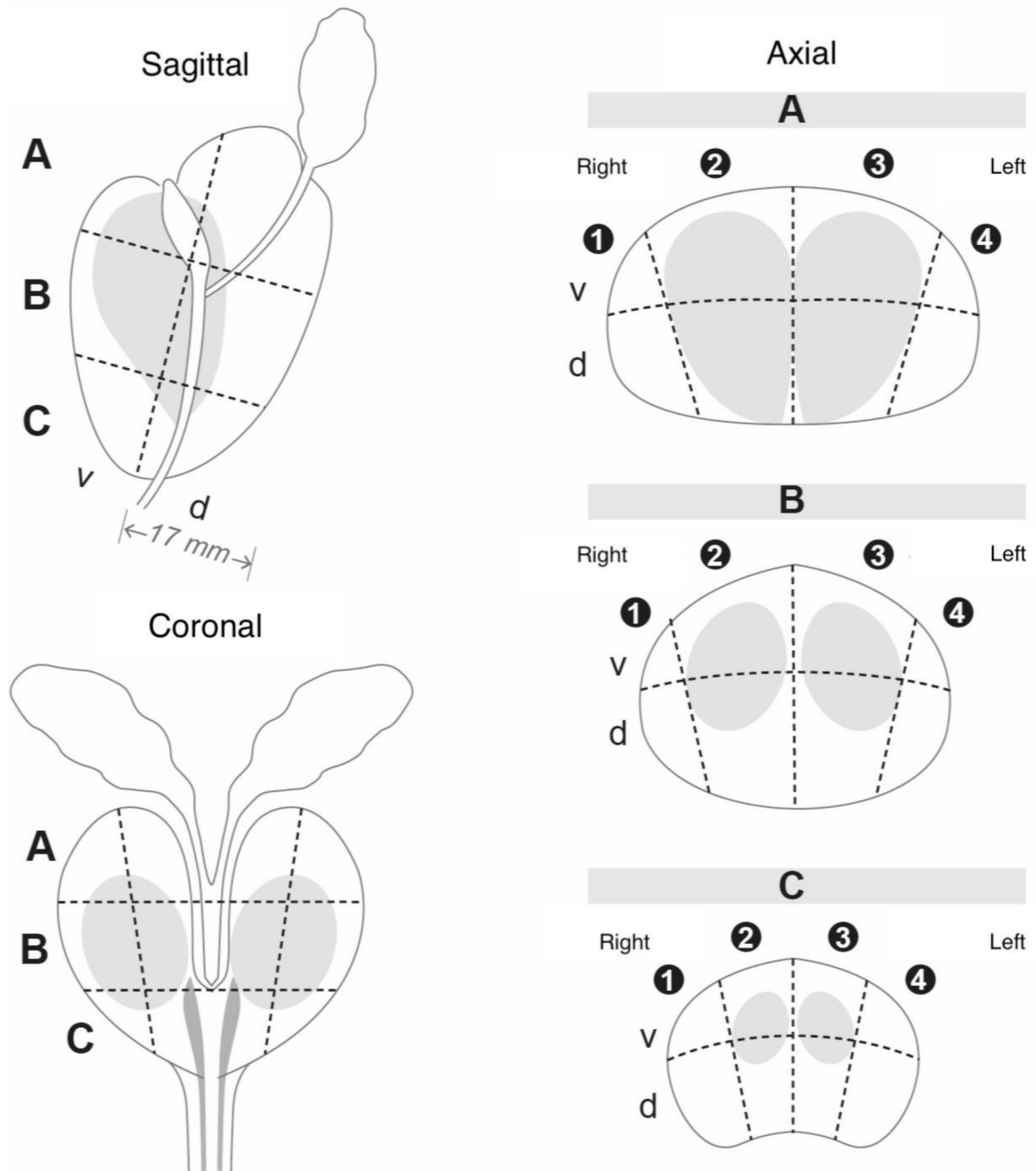


5. Soos G, Tsakiris I, Szanto J, et al. The prevalence of prostate carcinoma and its precursor in Hungary: an autopsy study. *Eur Urol*. 2005 11;48(5):739–44. doi: S0302-2838(05)00544-0[pii] 10.1016/j.eururo.2005.08.010. PubMed PMID: ; eng. [PubMed: 16203079]
6. Telesca D, Etzioni R, Gulati R. Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends. *Biometrics*. 2008 3;64(1):10–9. doi: BIOM825 [pii] 10.1111/j.1541-0420.2007.00825.x. PubMed PMID: ; eng. [PubMed: 17501937]
7. Loeb S, Bjurlin MA, Nicholson J, et al. Overdiagnosis and Overtreatment of Prostate Cancer [Review]. *Eur Urol*. 2014 6;65(6):1046–1055. doi: 10.1016/j.eururo.2013.12.062. PubMed PMID: ; Eng. [PubMed: 24439788]
8. Vårdprogramgruppen för prostatacancer Regionala cancercentrum i samverkan. Nationellt vårdprogram prostatacancer 2018, December 11 [2019-03–08]. Available from: <https://www.cancercentrum.se/samverkan/cancerdiagnoser/prostata/vardprogram/gallande-vardprogram-prostatacancer/>
9. Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet*. 2017 2 25;389(10071):815–822. doi: 10.1016/S0140-6736(16)32401-1. PubMed PMID: . [PubMed: 28110982]
10. Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. *N Engl J Med*. 2018 5 10;378(19):1767–1777. doi: 10.1056/NEJMoa1801993. PubMed PMID: . [PubMed: 29552975]
11. Sonn GA, Fan RE, Ghanouni P, et al. Prostate Magnetic Resonance Imaging Interpretation Varies Substantially Across Radiologists. *Eur Urol Focus*. 2017 12 6. doi: 10.1016/j.euf.2017.11.010. PubMed PMID: . [PubMed: 29226826]
12. Futterer JJ, Briganti A, De Visschere P, et al. Can Clinically Significant Prostate Cancer Be Detected with Multiparametric Magnetic Resonance Imaging? A Systematic Review of the Literature. *Eur Urol*. 2015 12;68(6):1045–53. doi: 10.1016/j.eururo.2015.01.013. PubMed PMID: ; eng. [PubMed: 25656808]
13. Moldovan PC, Van den Broeck T, Sylvester R, et al. What Is the Negative Predictive Value of Multiparametric Magnetic Resonance Imaging in Excluding Prostate Cancer at Biopsy? A Systematic Review and Meta-analysis from the European Association of Urology Prostate Cancer Guidelines Panel. *Eur Urol*. 2017 8;72(2):250–266. doi: 10.1016/j.eururo.2017.02.026. PubMed PMID: . [PubMed: 28336078]
14. Mottet N, Bellmunt J, Bolla M, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. *Eur Urol*. 2017 4;71(4):618–629. doi: 10.1016/j.eururo.2016.08.003. PubMed PMID: . [PubMed: 27568654]
15. Mottet N (Chair) RCNvdB, Briers E (Patient Representative), Cornford P (Vice-chair), De Santis M, Fanti S, Gillessen S, Grummet J, Henry AM, Lam TB, Mason MD, van der Kwast TH, van der Poel HG, Rouvière O, Tilki D, Wiegel T, Guidelines Associates: Van den Broeck T, Fossati N, Gross T, Lardas M, Liew M, Moris L, Schoots IG, Willemse P-PM. EAU - EANM - ESTRO - ESUR - SIOG Guidelines on Prostate Cancer: EAU Guidelines Office, Arnhem, The Netherlands; 2019 [28 6 2019]. Available from: <https://uroweb.org/guideline/prostate-cancer/>
16. Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. *Eur Urol*. 2016 1;69(1):16–40. doi: 10.1016/j.eururo.2015.08.052. PubMed PMID: . [PubMed: 26427566]
17. Mertan FV, Greer MD, Shih JH, et al. Prospective Evaluation of the Prostate Imaging Reporting and Data System Version 2 for Prostate Cancer Detection. *J Urol*. 2016 9;196(3):690–6. doi: 10.1016/j.juro.2016.04.057. PubMed PMID: . [PubMed: 27101772]
18. Seo JW, Shin SJ, Taik Oh Y, et al. PI-RADS Version 2: Detection of Clinically Significant Cancer in Patients With Biopsy Gleason Score 6 Prostate Cancer. *AJR Am J Roentgenol*. 2017 7;209(1):W1–W9. doi: 10.2214/AJR.16.16981. PubMed PMID: . [PubMed: 28418690]
19. Rosenkrantz AB, Ginocchio LA, Cornfeld D, et al. Interobserver Reproducibility of the PI-RADS Version 2 Lexicon: A Multicenter Study of Six Experienced Prostate Radiologists. *Radiology*. 2016 9;280(3):793–804. doi: 10.1148/radiol.2016152542. PubMed PMID: ; [PubMed: 27035179]

20. Kasel-Seibert M, Lehmann T, Aschenbach R, et al. Assessment of PI-RADS v2 for the Detection of Prostate Cancer. *European journal of radiology*. 2016 4;85(4):726–31. doi: 10.1016/j.ejrad.2016.01.011. PubMed PMID: . [PubMed: 26971415]
21. Muller BG, Shih JH, Sankineni S, et al. Prostate Cancer: Interobserver Agreement and Accuracy with the Revised Prostate Imaging Reporting and Data System at Multiparametric MR Imaging. *Radiology*. 2015 12;277(3):741–50. doi: 10.1148/radiol.2015142818. PubMed PMID: ; [PubMed: 26098458]
22. Cohen J A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*. 1960;20(1):37–46.
23. Greer MD, Brown AM, Shih JH, et al. Accuracy and agreement of PIRADSV2 for prostate cancer mpMRI: A multireader study. *J Magn Reson Imaging*. 2017 2;45(2):579–585. doi: 10.1002/jmri.25372. PubMed PMID: . [PubMed: 27391860]
24. de Rooij M, Hamoen EH, Futterer JJ, et al. Accuracy of multiparametric MRI for prostate cancer detection: a meta-analysis. *AJR Am J Roentgenol*. 2014 2;202(2):343–51. doi: 10.2214/AJR.13.11046. PubMed PMID: . [PubMed: 24450675]
25. Borofsky S, George AK, Gaur S, et al. What Are We Missing? False-Negative Cancers at Multiparametric MR Imaging of the Prostate. *Radiology*. 2018 1;286(1):186–195. doi: 10.1148/radiol.2017152877. PubMed PMID: ; [PubMed: 29053402]
26. Rosenkrantz AB, Ayoola A, Hoffman D, et al. The Learning Curve in Prostate MRI Interpretation: Self-Directed Learning Versus Continual Reader Feedback. *AJR Am J Roentgenol*. 2017 3;208(3):W92–W100. doi: 10.2214/AJR.16.16876. PubMed PMID: . [PubMed: 28026201]
27. Niu XK, Chen XH, Chen ZF, et al. Diagnostic Performance of Biparametric MRI for Detection of Prostate Cancer: A Systematic Review and Meta-Analysis. *AJR Am J Roentgenol*. 2018 8;211(2):369–378. doi: 10.2214/AJR.17.18946. PubMed PMID: . [PubMed: 29894216]
28. Kang Z, Min X, Weinreb J, et al. Abbreviated Biparametric Versus Standard Multiparametric MRI for Diagnosis of Prostate Cancer: A Systematic Review and Meta-Analysis. *AJR Am J Roentgenol*. 2018 12 4;W1–W9. doi: 10.2214/AJR.18.20103. PubMed PMID: . [PubMed: 30512996]



**Figure 1.** Flow chart of the patients included in the study.



**Figure 2.**

The 24-sector template, based on the Swedish National PC Guidelines, used to record the anatomical localization of the MRI-lesions and the tumour localization on the whole mount sections. The index tumours were considered detected if there was a lesion with a PI-RADS score 3–5 described in the same or neighbouring sector (approximate match). A = base, B = mid-gland, C = apex v = ventral, d = dorsal

**Table 1.**

Patient and tumour characteristics, both pre- and postoperatively. Numbers represent frequency (%) or median (inter quartile range).

<b>Preoperative characteristic</b>		
<b>Age at surgery, years</b>	61	(56; 65)
<b>PSA at diagnosis, ng/mL</b>	6.4	(4.3; 9.7)
<b>TRUS-estimated prostate volume, cc</b>	31	(25; 42)
<i>Missing n=4</i>		
<b>PSA density</b>	0.21	(0.13; 0.35)
<i>Missing n=4</i>		
<b>Total number of biopsy cores taken</b>	10	(10; 12)
<i>Missing n=2</i>		
<b>Number of positive biopsy cores</b>	4	(2; 6)
<i>Missing n=4</i>		
<b>Percentage positive biopsy cores (%)</b>	40%	(20%; 65%)
<i>Missing n=4</i>		
<b>Total cancer core length, mm</b>	13	(6; 25)
<i>Missing n=3</i>		
<b>Clinical stage</b>		
T1c	69	(71.1%)
T2	26	(26.8%)
T3-4	0	(0%)
<i>Missing</i>	2	(2.1%)
<b>Biopsy Gleason score</b>		
3+3	41	(42.3%)
3+4	36	(37.1%)
4+3	11	(11.3%)
4+4	9	(9.3%)
<b>Postoperative outcomes</b>		
	<b>n</b>	<b>(%)</b>
<b>Operation technique</b>		
Retropubic radical prostatectomy	84	(86.6%)
Robot-assisted laparoscopic prostatectomy	13	(13.4%)
<b>Pathological stage</b>		
pT2	61	(62.9%)
pT3	36	(37.1%)
<b>Pathological Gleason score</b>		
<i>Small/insignificant 3+3 (&lt; 10 mm)</i>	5	(5.2%)
<i>3+3 ( 10 millimetre or unknown size)</i>	18	(18.6%)
<i>3+4</i>	48	(49.5%)
<i>4+3</i>	17	(17.5%)

---

---

<b>Preoperative characteristic</b>		
<i>4+4</i>	9	(9.3%)
<b>Tumour diameter &gt; 10 millimetre</b>		
Yes	75	(77.3%)
No	5	(5.2%)
<i>Missing</i>	17	(17.5%)
<b>Surgical margins</b>		
Positive	18	(18.6%)
Negative	76	(78.4%)

---

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2a.**

Index tumour reported as PI-RADS 3–5 by the three readers separately and average index tumour detection rate. Numbers represents percent (n).

Gleason score (GS)	Reader 1 (200 reports before review) Missing = 1	Reader 2 (50 reports before review) Missing = 1	Reader 3 (300 reports before review) Missing = 0	Average detection rate
3+3 (< 10 mm) n=5	40% (2)	0% (0)	40% (2)	26.7%
3+3 ( 10 mm /unknown size) n=18	83.3% (15)	58.8% (10) missing=1	72.2% (13)	70.4%
3+4 n=48	70.8% (34)	70.8% (34)	77.1% (37)	72.9%
4+3 n=17	88.2% (15)	70.6% (12)	82.4% (14)	80.4%
4+4 n=9	87.5% (7) missing=1	88.9% (8)	88.9% (8)	88.4%
All GS pT2 n=61	75.0% (45) missing=1	61.7% (37) missing=1	73.8% (45)	70.2%
All GS pT3 n=36	77.8% (28)	75.0% (27)	80.6% (29)	77.8%
Total n=97	76.0% (73)	66.7% (64)	76.3% (74)	73.0%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2b.**

Index tumour reported as PI-RADS 4–5 by the three readers separately and average index tumour detection rate. Numbers represents percent (n).

Gleason score (GS)	Reader 1 (200 reports before review) Missing=1	Reader 2 (50 reports before review) Missing=1	Reader 3 (300 reports before review) Missing=0	Average detection rate
3+3 (< 10 mm)n=5	20.0% (1)	0% (0)	0% (0)	6.7%
3+3 ( 10 mm /unknown size) n=18	66.7% (12)	29.4% (5) missing=1	50.0% (9)	48.7%
3+4 n=48	54.2% (26)	60.4% (29)	68.8% (33)	61.1%
4+3 n=17	82.4% (14)	64.7% (11)	82.4% (14)	76.5%
4+4 n=9	87.5% (7) missing=1	77.8% (7)	88.9% (8)	84.7%
All GS pT2 n=61	55.0% (33) missing=1	48.3% (29) missing=1	60.7% (37)	54.7%
All GS pT3 n=36	75.0% (27)	63.9% (23)	75.0% (27)	71.3%
Total n=97	62.5% (60)	54.2% (52)	66.0% (64)	60.9%

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Table 3.**

Agreement for pairwise combination of readers with lesions PI-RADS 3 and PI-RADS 4. Numbers represents Cohen's  $\kappa$  coefficient (95% confidence interval).

PI-RADS v.2.limit	Reader 1 vs. 2	Reader 1 vs. 3	Reader 2. vs. 3
At least PI-RADS 3	0.38 (0.18,0.58)	0.31 (0.09,0.53)	0.54 (0.35,0.72)
At least PI-RADS 4	0.50 (0.32,0.67)	0.49 (0.31,0.68)	0.54 (0.37,0.71)

**Table 4.**

Characteristics of index tumours with negative preoperative MRI.

**Table 4a** – Characteristics of index tumours that no reader reported on MRI.

Age at surgery (years)	PSA (ng/mL)	Biopsy GS	Pathological stage	Pathological GS	Localization of index tumour	Size of index tumour(mm)	SVI	ECE	PSM	Tumour 2 or 3 detected *
71	8	3+3	pT2c	3+4	4Cvd	10	No	No	No	
66	4	3+4	pT2c	3+4	1Cd	14×9	No	No	No	
67	5	3+3	pT2c	3+4	12BCvd	19×16	No	No	No	
65	8	3+4	pT3a	3+4	1234Cv	22×11	No	Yes	Yes	Yes, GS 3+4
64	3	3+4	pT3a	3+4	1Avd	13×4	No	Yes	No	
63	22	3+4	pT3b	4+5	1Ad	Unknown	Yes	Yes	Yes	Yes, GS 3+4
61	6	3+3	pT2a	3+3	1Cd	15×4	No	No	No	
59	4	3+3	pT2c	3+3	1Cd	8×5	No	No	No	
56	7	3+4	pT3a	3+4	4Bvd	16×8	No	Yes	No	Yes, GS 3+4
56	4	3+3	pT2	4+3	4ABCv	18×8	No	No	No	
52	3	3+3	pT2a	3+3	1Avd	2	No	No	No	

\*One or more of the readers described a lesion where the second or third tumour on the specimen were found

**Table 4b**

– Characteristics of index tumours that only one of three reader reported on MRI.

Age at surgery (years)	PSA (ng/mL)	Biopsy GS	Pathological stage	Pathological GS	Localization of index tumour	Size of index tumour (mm)	SVI	ECE	PSM
73	5	3+4	pT2b	3+4	1ABd	16×8	No	No	No
71	10	4+5	pT2c	4+3	4ABCvd	Unknown	No	No	Yes
65	4	3+3	pT2c	3+4	12Bd	Unknown	No	No	No
62	4	3+3	pT2c	3+4	1Avd	10×7	No	No	No
64	4	3+4	pT2c	3+4	34BCd	20×8	No	No	No
62	4	3+3	pT2c	3+3	4ABCvd	3×3	No	No	No
60	4	3+3	pT3a	3+4	34BCvd	Unknown	Yes	Yes	No
59	6	3+3	pT3a	3+3	4BCd	Unknown	No	Yes	Yes
59	4	3+3	pT2a	3+3	34Bv	19×9	No	No	No
58	4	3+3	pT2	3+3	2Bd	9×5	No	No	No
45	30	3+4	pT3b	4+3	1234ABCvd	Unknown	Yes	Yes	Yes

GS = Gleason score, SVI = Seminal vesicle invasion, ECE = Extracapsular extension, PSM = Positive Surgical margins