UNIVERSITA' VITA-SALUTE SAN RAFFAELE

CORSO DI DOTTORATO DI RICERCA INTERNAZIONALE IN MEDICINA MOLECOLARE

Curriculum in Medicina Clinica e Sperimentale

Simplification of prostate MRI protocol and development of a novel MRI technique for prostate cancer detection and characterization

DoS: Prof. Francesco De Cobelli

Second Supervisor: Prof. Shonit Punwani

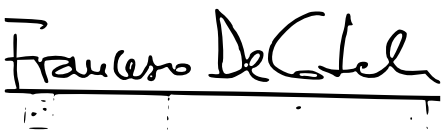
Tesi di DOTTORATO di RICERCA di Giorgio Brembilla

matr. 013871

Ciclo di dottorato 2018/2019

SSD MED/36

Anno Accademico 2020/2021



CONSULTAZIONE TESI DI DOTTORATO DI RICERCA

Il sottoscritto / I **Giorgio Brembilla**
Matricola / *registration number* **013871**
nato a/ *born at* **Ponte San Pietro (BG)**
il/*on* **22/09/1987**

autore della tesi di Dottorato di ricerca dal titolo / *author of the PhD Thesis titled*

Simplification of prostate MRI protocol and development of a novel MRI technique for prostate cancer detection and characterization

NON AUTORIZZA la Consultazione della tesi /**DOES NOT AUTHORIZE** *the public release of the thesis*

a partire dalla data di conseguimento del titolo */ from the PhD thesis date*

Poiché /*because*:

□ l'intera ricerca o parti di essa sono potenzialmente soggette a brevettabilità/ *The whole project or part of it might be subject to patentability;*

□ ci sono parti di tesi che sono già state sottoposte a un editore o sono in attesa di pubblicazione/ *Parts of the thesis have been or are being submitted to a publisher or are in press;*

Si rende noto che parti della tesi sono indisponibili in relazione all'utilizzo di dati tutelati da segreto industriale /*Please Note: some parts of the thesis are not available in relation to the norm of the use of information protected by trade secret*

E' fatto divieto di riprodurre, in tutto o in parte, quanto in essa contenuto / *Copyright the contents of the thesis in whole or in part is forbidden*

Data /Date ..... Firma /Signature .....

DECLARATION

This thesis has been composed by myself and has not been used in any previous application for a degree. Throughout the text I use both 'I' and 'We' interchangeably.

All the results presented here were obtained by myself, except for:

- LI-MRI quantitative values (chapter "*Shortened Luminal Index MR imaging (LI-MRI) for the detection and characterization of significant PCa*") were derived by a software created by Ms Fiona Gong at the UCL Centre for Medical Imaging (2nd Floor Charles Bell House, 43-45 Foley Street, W1W 7TS, London, UK), under the supervision of Prof. Shonit Punwani and Prof. David Atkinson. The pipeline for Luminal Water Imaging acquisition and processing is patented by UCL and is protected by intellectual property.

The following chapter contain data that have been already published:

- **Interreader Agreement of PI-RADS v2.1**: Brembilla et al. 2020, DOI: https://doi.org/10.1007/s00330-019-06654-2
- **Tackling interobserver variability in multiparametric MRI**: Brembilla et al. 2021, DOI: https://doi.org/10.1016/j.eururo.2020.10.023
- **Diagnostic accuracy of abbreviated bi-parametric MRI (a-bpMRI) for prostate cancer detection and screening**: Brembilla et al. 2022, DOI: https://doi.org/10.3390/ diagnostics12020231

All sources of information are acknowledged by means of reference.

Abstract

Background & rationale: Multiparametric MRI (mpMRI) is the imaging modality of choice for the diagnostic workup of men with clinically suspected prostate cancer. However, mpMRI has several limitations, including: limited accessibility in routine clinical practice; high interreader variability; low specificity and positive predictive value.

Aim of the work: The aim of the work is to address current limitations of mpMRI with focus on MRI protocol simplification, interreader agreement assessment and development of a novel MRI technique called Luminal Index MRI (LI-MRI).

Material & Methods: I) The interreader variability of mpMRI was investigated in a multireader study reflecting the real-world clinical practice in 200 patients. Seven radiologists reviewed and scored the mpMRI scans using PI-RADS v2.1. Agreement on index lesion detection was evaluated with Conger's k coefficient, agreement coefficient 1 (AC1), percentage of agreement (PA) and indexes of specific agreement. II) The feasibility of MRI protocol simplification was tested on 151 men who underwent mpMRI and transperineal template prostate mapping biopsies. Three experienced radiologists scored MRI scans using three different protocols (mpMRI, bpMRI and abbreviated bp-MRI [a-bpMRI]). The diagnostic performance and interreader agreement of the different protocols was compared. III) LI-MRI diagnostic performance was tested on a total of 178 men who underwent mpMRI for clinically suspected PCa. A validation cohort was then used to validate the results and to test LI-MRI reproducibility. IV) Retrospective analysis on men who underwent LI-MRI (+/-biopsy) was performed to develop a dedicated scoring system for LI-MRI.

Results: I) Agreement on index lesion detection among was substantial (AC1 0.738; 95% CI 0.695–0.782); dedicated radiologists had higher agreement compared with non-dedicated readers. II) The sensitivity and specificity of a-bpMRI were 92% and 48%, respectively. There was no significant difference in sensitivity compared to bpMRI and mpMRI. Interreader agreement of a-bpMRI was moderate (AC1 0.58). III) LI-MRI performed using a simplified 8-echo protocol was feasible, repeatable and yielded high specificity for the detection of clinically significant prostate cancer (65-78% specificity for 89-90% sensitivity). IV) A dedicated scoring system was developed for LI-MRI based on both multiecho-T2W and LI-maps.

Conclusion: The variability of mpMRI is lower than previously thought. Short MRI protocols may perform as well as mpMRI in expert hands and should be considered to address the limited accessibility of mpMRI. LI-MRI is a promising tool that could improve PCa characterization with MRI.

Table of Contents

Acronyms and abbreviations

mpMRI: multiparametric MRI

bpMRI: biparametric MRI

a-bpMRI: abbreviated biparametric MRI

LI-MRI: luminal index MRI

LWF: luminal water fraction

PCa: prostate cancer

csPCa: clinically significant prostate cancer

TRUS: transrectal ultrasound

Bx: biopsy

List of tables and figures

Introduction

Epidemiology

Prostate cancer (PCa) is the second most common tumour in men, accounting for more than 1 million cases worldwide in the 2020. Autopsy studies showed a PCa prevalence of more than 50% by age > 79 years (Bell *et al*, 2015). In Europe, the lifetime cumulative risk of PCa is approximately 16% while the all-age incidence is 63 per 100.000 males. PCa is the fifth cause of cancer-related death worldwide, being the first cancer-related death cause in approximately one third of the countries. In general incidence is higher and mortality is lower in populations with higher income and Human Development Index (HDI) (Gandaglia *et al*, 2021).

Numerous risk factors have been associated with increased lifetime risk of PCa, including (Gandaglia *et al*, 2021):

- Family history: the risk of PCa is increased for each first-degree relative who is/has been affected by PCa, especially if younger than sixty years old. In general, up to twenty percent of men with PCa have a family history of the disease. No single gene has been associated to familiar history, but some have been identified to have a potential role including HOXB13 (high risk), BRCA mutations and MSH2.
- Ethnicity and race (e.g., African origin) are well known risk factors of PCa and may partially explain geographic variations in incidence and mortality of PCa. In turn, the mechanism that underlies such variability may be explained by genetic predisposition. However, disparities in socio-economic factors linked to ethnicity might also play an important role.
- Lifestyle and environmental factors may influence PCa incidence and outcome. Among the others, metabolic syndrome has been associated with an increased risk of PCa, smoking and obesity with increased mortality; conversely, physical activity might improve PCa outcome. No conclusive evidence is available on the on the impact of diet and drugs on PCa incidence to make reliable recommendations.

A correct identification of people at higher risk of PCa incidence/progression/mortality has profound clinical implications on screening programs and diagnostic approach.

Anatomo-pathology

Histologic anatomy

Anatomically, the prostate is usually divided into three main zones: peripheral (PZ), transition (TZ) and central zone (CZ). The outer boundary of the prostate is enclosed in a pseudocapsule, made of condensed fibromuscular stroma (rather than a real capsule). The CZ lies between the ejaculatory ducts and the bladder neck and represents about one third of the prostatic glandular mass. Histologically, the central zone contains a distinct type of glands which show a more complex architecture compared other zones. The TZ surrounds the urethra from which is separated by dense fibromuscular stroma (pseudocapsule). The PZ normally represents the most part of the prostatic tissue and is predominantly located in the posterior aspect of the prostate but it extends also laterally and anteriorly (anterior horns).

About two-thirds of PCa originate in the PZ, mostly posteriorly or postero-laterally. The second most common location is the TZ, which harbours usually higher volume tumours even if with a comparable outcome to those arising in the PZ. CZ tumours are rarer.

The prostatic gland is composed of acini and ducts, lined by an inner layer of luminal cells and an outer layer of basal cells, interspersed in a fibromuscular stroma. Only the luminal cells express PSA, while androgen receptors can be found also in stromal cells. Cancer is associated with a proliferation of the cellular component with reduction of glandular lumen and stromal component [].

Age-related changes occur in each zone. Typically, changes in the TZ consist in hyperplasia of both the glandular and fibromuscular components (benign prostatic hyperplasia - BPH). Conversely, atrophic changes typically occur in the PZ with different morphological patterns (e.g., simple, cystic, sclerotic and hyperplastic) together with acute/chronic inflammatory changes.

Prostate carcinoma subtypes

It is commonly accepted that some histological changes in prostate glands may represent cancer precursors, namely high grade prostatic intraepithelial neoplasia (HG-PIN). HG-PIN is characterized by dysplastic luminal cells in acini or ducts, in the context of a retained glandular architecture. The histological hallmark are prominent nucleoli in luminal cells. HG-PIN is frequently found in coexistence with invasive carcinoma, with which shares similar cytologic and genetic changes. However, its progression to invasive PCa is still under debate.

Acinar adenocarcinoma represents the most common subtype of PCa (>95%). It is characterized by high heterogeneity, with the formation of small sized glands that can fuse or form alternative architectural patterns (e.g., cribriform pattern that is associated with a worse prognosis). Several variants of acinar adenocarcinoma exist

(e.g., foamy gland, pseudo hyperplastic, mycrocistic, signet ring cell, pleomorphic giant cell, sarcomatoid, mucinous).

Ductal adenocarcinoma accounts for approximately 1% of the cases as an isolated entity, but some consider it as a variant of acinar adenocarcinoma and they more frequently occur in association. Ductal adenocarcinoma typically involves large periurethral ducts with distinct cellular changes compare to the acinar variant. Ductal adenocarcinoma is more aggressive than the acinar subtype and is generally categorised as Gleason pattern 4-5.

The presence of neuroendocrine cells in acinar adenocarcinoma is not infrequent; however, poorly differentiated neuroendocrine carcinomas exist either small or large cell patterns. They rarely occur in their pure form (in case of which a metastasis from other sites should be hypothesized) and have an aggressive behaviour.

Other rare subtypes include basal, squamous and adenosquamous cell carcinoma.

Prostate cancer grading

Prostate cancer is graded using the Gleason grading system that evaluates morphological and architectural features of PCa assigning a score from 1 (well differentiated) to 5 (poorly differentiated). It accounts for the heterogeneity of PCa by assigning primary and secondary scores, obtaining a scale that ranges from 2 to 10 (e.g., Gleason 3+4). In practice, scores 1 and 2 are no longer used resulting in a possible range of 6-10. Differences exist when assigning Gleason scores in biopsy vs prostatectomy specimens. In biopsy specimens the score is calculated by adding the most dominant pattern and the highest grade of the remaining patterns. In prostatectomy specimens the score is obtained scoring the first and second dominant patterns.

More recently, the International Society of Urological Pathology (ISUP) has adopted a new grading system that includes only 5 Gleason grade groups, to allow a simplified interpretation and yield a better stratification based on aggressiveness and outcome (Epstein *et al*, 2016). Also, it helps patients understand better the grade of the disease (e.g., low grade for non-harmful tumour is Grade 1/5 rather than Gleason 3+3).

Nevertheless, this classification suffers from high variability in reporting, especially for Gleason pattern 4, and additional information may be useful to better stratify tumour aggressiveness and prognosis (Wolters *et al*, 2010). Among the others, the quantification (%) of high grade pattern (4-5) on the total is encouraged as it is associated with recurrence and represents an independent prognostic factor (Vis *et al*, 2007). Also, pattern 4 is the most heterogeneous and may include several

different morphological patterns (cribriform, glomeruloid, fused glands and poorly formed glands) that have different clinical behaviour. Among them, it is recognized that the cribriform pattern is associated with worse outcome and its presence should be mentioned in any histopathological report (Kweldam *et al*, 2015). Similarly, the presence of intraductal carcinoma (IDC-P) is often associated with high grade invasive carcinoma and should be reported as well (Guo & Epstein, 2006).

PCa is frequently multifocal at histopathological analyses, with 80% of the prostatectomy specimens finding 2-5 lesions.

Diagnostic evaluation of PCa

Screening & early detection

Population screening consists in a large-scale, systematic examination of a whole asymptomatic population at risk (of PCa). It is intended to yield an early detection of the disease, resulting in a reduction of mortality/morbidity related to the disease and a maintained quality of life (QoL). Despite its popularity, PCa screening is still one of the most debated and controversial topics in the urological community. Still, no consensus exists on the use of PSA testing for population-based screening, since results from large screening studies has been conflicting; also, national guidelines and recommendations are inconsistent (Mottet *et al*, 2020). Concerns over PCa screening are linked to the high prevalence of latent disease in the general population, with high prevalence of PCa in the healthy population seen in autopsy studies (Bell *et al*, 2015). Population-based screening can accidentally lead to the detection of such latent (i.e. non-harmful) disease, causing overdiagnosis, overtreatment and consequent harm to the patient.

The ideal screening tool should allow an early diagnosis of harmful disease to prevent disease-related morbidity and mortality. PCa is characterized by a prolonged natural history, and a clear definition of what represents clinically significant cancer (csPCa) is unclear. For example, Gleason score ≥3+4 (ISUP grade group ≥2) is the most common definition of csPCa in the literature; however, long term cancer-specific mortality of ISUP grade group 2 disease may be similar to that of non-significant PCa (nsPCA; Gleason 3+3, ISUP grade group 1), with a significant increased risk of mortality only for Gleason score of ≥4+3 disease (ISUP grade group ≥3) (Bill-Axelson *et al*, 2018).

Limitations of PSA. There is currently no consensus on the optimal PSA value to be used for screening. The main limiting factor is the low diagnostic accuracy of PSA. For example, using a threshold of 3 ng/ml (as shown in the prostate cancer prevention trial – PCPT), the sensitivity of PSA for ISUP grade group ≥2 is 58%, while a threshold of 1.6 ng/ml would be needed to reach a sensitivity of 84% (Thompson *et al*, 2005). Also, specificity is low due to several common benign conditions that can cause PSA elevation (e.g. benign hyperplasia, infection/inflammation).

Effect on mortality. Several randomized controlled trials failed to show any significant impact of PSA-based screening on mortality. However, variability in study design and methodological limitations may limit the interpretability of the results. Conversely, the European Randomized study of Screening for Prostate Cancer (ERSPC) has been shown to have a low risk of bias. The ERSPC study reported a 20% reduction of mortality at 16 years (Hugosson *et al*, 2019). However, the number needed to screen (NNS) is still high (570 men) compared to other screening modalities such as mammography (NNS = 235).

Benefit to harm ratio. Transrectal ultrasonography (TRUS) guided biopsy is historically used to confirm the diagnosis of PCa in men with raised PSA within screening programs. However, it is known to be associated with serious adverse events (e.g. infections) and significant morbidity (hospitalization rates of up to >1%). Roughly one third of men experience moderate or major adverse events (Rosario *et al*, 2012). However, the most serious side effect of PSA screening is overdiagnosis, that is estimated to occur in up to 40%. Overdiagnosis leads carries a significant psychological burden of cancer diagnosis and, more important, may expose men to severe side effects and morbidity related to radical treatments (overtreatment).

PSA measurement (and eventually DRE) need to be repeated, but no consensus exists on the optimal intervals for testing. A strategy based on the initial PSA – derived risk estimation has been proposed: every 2 years for men at risk; postponed up to 8 years in men not at risk (initial PSA < 1 ng/mL at 40 years; PSA < 2 ng/mL at 60; no familiar history) (Gelfond *et al*, 2015).

Even an individualized risk-adapted strategy is associated with a high risk of overdiagnosis. Consequently, the dissociation of the link between diagnosis and treatment is key to reduce the harms of overtreatment.

Clinical Diagnosis

Clinical diagnosis of PCa is generally based on PSA levels and/or DRE. Definitive diagnosis is made on histopathological analysis of prostate biopsy samples.

DRE. The majority of PCa are located in the peripheral zone (PZ) and can be detected by DRE when they are >0.2 cc. An abnormal DRE is associated with an increased risk of aggressive PCa and is an indication for biopsy.

PSA. It is an organ specific serum marker (not cancer specific) that can be elevated also in some benign conditions such as IPB and inflammation. No single threshold exist that is related to the risk of harbouring PCa, rather the risk increases with increasing values of PSA (Mottet *et al*, 2020) (Table 1). PSA density is the value of PSA divided by the volume, and is associated with csPCa. PSA velocity (absolute annual increase) and PSA doubling time may have a prognostic role but have a limited diagnostic role. The free/total PSA ratio can be used to stratify the risk of PCA for PSA values 4-10 ng/ml.

PSA level (ng/mL)	Risk of PCa (%)	Risk of ISUP $\geq$ 2 (%)
0.0–0.5	6.6	0.8
0.6–1.0	10.1	1.0
1.1–2.0	17.0	2.0
2.1–3.0	23.9	4.6
3.1–4.0	26.9	6.7

Table 1. *Risk of PCA and PSA values (Thompson et al, 2004)*

Risk calculators. Risk calculators may help determining the risk of PCa to avoid unnecessary biopsies. Among the others available:

- the PCPT cohort: PCPTRC 2.0 http://myprostatecancerrisk.com/;
- the ERSPC cohort: http://www.prostatecancer-riskcalculator.com/seven-prostate-cancer-risk-calculators;

The role of Imaging for PCa diagnosis

Transrectal Ultrasound (TRUS) has been historically used to evaluate prostate volume and morphology, and to eventually identify suspicious area for PCa. However, TRUS is not accurate in detecting PCa both in the PZ and the TZ (Rouvière *et al*,

2019). Emerging applications of TRUS have been investigated, including sonoelastography, micro-Doppler, contrast-enhanced US, especially in combination ("multiparametrc US"), however no evidence is available to support their use in clinical practice. Promising results have been described for high resolution micro-ultrasound probes, but this technique requires validation in large cohorts. Among the major drawbacks of advanced TRUS techniques are the limited standardizability and high variability in the acquisition/interpretation of US findings.

CT does not have sufficient soft tissue contrast to characterize prostate gland.

In recent years, multiparametric MRI has established itself as the imaging modality of choice in the diagnostic workup of prostate cancer, following the results of large prospective trials (Mottet *et al*, 2020). Thanks to its ability to discriminate normal vs pathological prostatic tissue, RM has high sensitivity (>90%) for the detection of csPCa, and allows to reduce the number of unnecessary prostate biopsies that would be performed based on clinical suspicion (Kasivisvanathan *et al*, 2018). Prostate MRI technique will be discussed in details in the following chapters.

Multiparametric MRI (mp-MRI) of the prostate

Mp-MRI is the modality of choice for the diagnostic workup of PCa. The term multiparametric is referred to the standard protocol of MRI acquisition for the evaluation of the prostate that includes: T2-weighted imaging (T2W), diffusion-weighted imaging (DWI) with the reconstruction of apparent coefficient of diffusion (ADC) maps, dynamic contrast-enhanced imaging (DCE).

Normal anatomy of the prostate on MRI

From cranial to caudal location, the prostate is divided into the base, the midgland, and the apex. Thanks to its high soft tissue contrast resolution, T2-weighted (T2W) MRI imaging allows a clear distinction of the different histologic zones:

- the peripheral zone: its appearance varies according to patient's age. In a normal sized prostate, it represents up two-thirds of the entire gland volume. It has a bright signal on T2W images due to the predominance of glandular lumens, it does not show significant restriction of diffusion on DWI sequences (and is bright on ADC maps), and it has only faint enhancement. Its typical signal is related to the cystic degeneration of prostate glands with age: in young men, the signal of the PZ may be less bright on T2, with more pronounced background enhancement, due to a more represented cellular component of the glands. Approximately 75% of prostate cancers occur in the PZ.

- The transition zone: it is composed of two symmetric lobes on each side of the prostatic urethra from base to midgland level, while it is usually absent at the apex. It has a different histological composition compared to the PZ, hence different signal characteristics at MRI. While it only represents 5% of the gland volume in young adults, it grows with age due to benign prostatic hyperplasia (BPH) and can compress/displace the normal PZ. On T2W images, TZ typically shows high heterogeneity. In young adults , it shows uniformly low signal intensity, while in the presence of hyperplasia it forms encapsulated nodules with variable composition (glandular: high T2 signal; stromal: low T2 signal), with low signal areas between nodules stroma. The TZ boundaries are usually well demarcated by a pseudo-capsule (also called surgical capsule) that is composed of compressed fibro-muscular tissue and divides the TZ from the PZ.
- The central zone: it is a symmetrical conical-shaped area that sorrounds the ejaculatory ducts

Sector map. To facilitate the communication between radiologists, urologists and pathologists, the use of a standardized sector map has been proposed by a European Consensus Meeting and the ESUR Prostate MRI Guidelines 2012 (Barentsz *et al*, 2012). It includes 41 sectors/regions including the prostate, the seminal vesicles and the external urethral sphincter. The aim of the sector map is to standardize reporting and allow a more precise localization of benign/pathological findings biopsy, therapy (focal therapies, active surveillance, surgical planning), pathological correlation, and research.

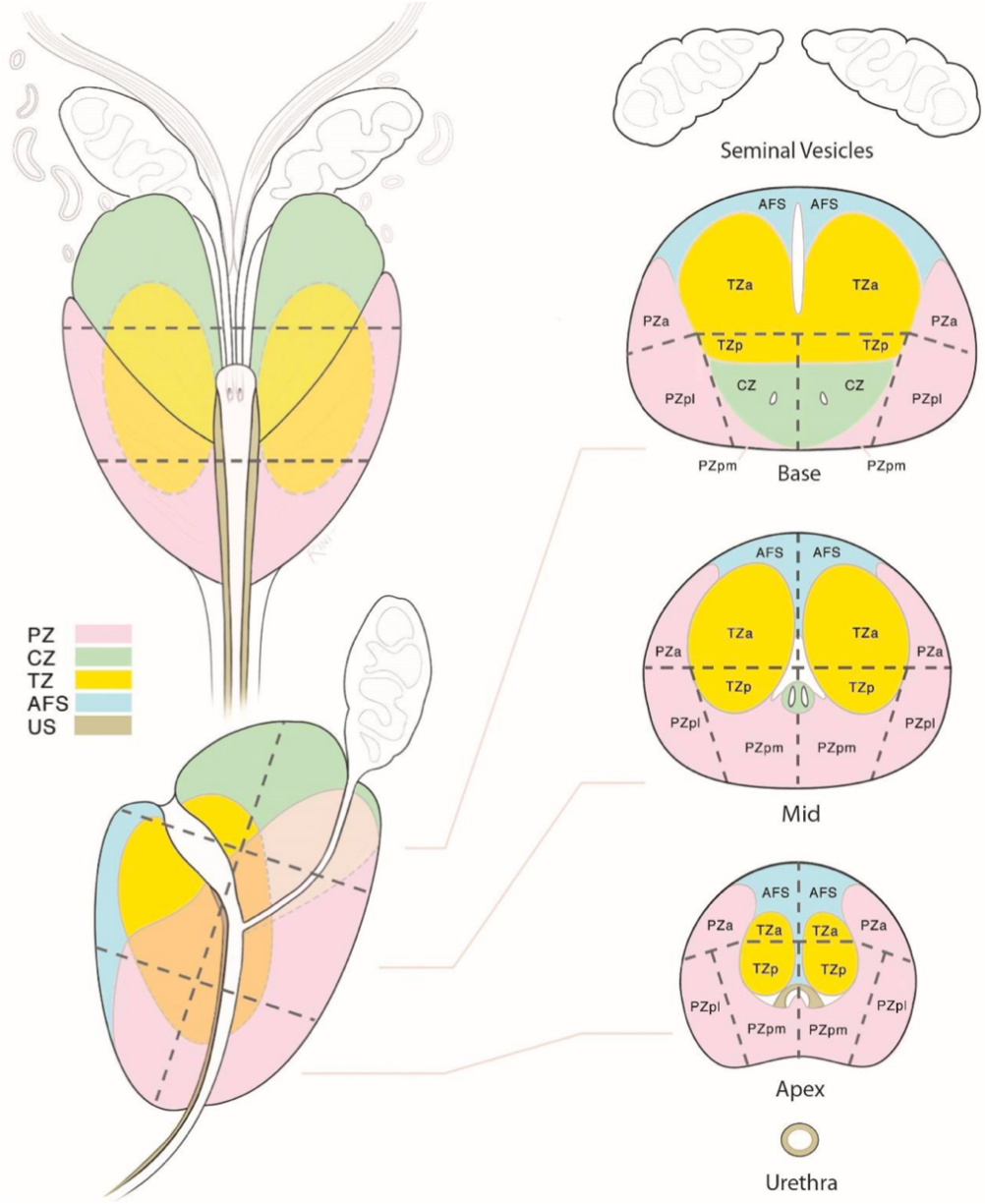


Figure 1. *Sector map of the prostate* (Turkbey *et al*, 2019).

Benign findings. It is very common to encounter MRI signal changes in the normal prostate (both in the PZ and TZ) that represent benign pathology. Among the most common:

- *Benign prostatic hyperplasia (BPH).* It develops in response to the physiological hormonal stimulation (androgens) of the TZ, although ectopic/exophytic BPH can be found in the PZ/CZ as well. BPH consists in a mix of stromal/glandular hyperplasia forming band-like areas and encapsulated round nodules. Glandular BPH nodules and cystic are characterized by hgh T2 signal, while stromal BPH nodules are are

characterized by low T2 signal. More commonly, BPH nodules are composed of a mixture of signal intensities. Depending on their composition, BPH nodules may show hypervascularity on DCE and have variable signal intensities on DWI. Of note, the stromal component can have similar MRI signal characteristics compared to PCa, causing difficulties in the differential diagnosis with potential false-positive findings at MRI. Per se, BPH is a benign process, but it can be associate with urinary obstructions symptoms with significant clinical implications. Also, BPH produces PSA, increasing its blood levels even in the absence of PCa.

- *Prostatitis.* Prostatitis is an inflammatory process characterized by a variable immune infiltrate depending on the causative agent. On MRI, it is associated with several signal changes in the PZ (and TZ), typically: hypointensity on T2 and the ADC map, variably reduced diffusion on DWI, hypervascularity on DCE. These characteristics can overlap significantly with those of PCa, potentially causing false-positive findings. At the same time, diffuse signal changes may mask underlying micro-foci of PCa or low-grade PCa. The main distinctive characteristic (compared to PCa) is the morphology that is commonly band-like/wedge-shaped, or diffuse rather than focal, and the restriction of the diffusion is usually not as marked/focal as in PCa.
- *Atrophy.* Prostatic glandular atrophy commonly occur as a normal aging process or can result from chronic inflammation. On MRI, the loss of glandular tissue is associated with several signal changes, typically: wedge-shaped hypointense areas on T2W, mildly reduction of diffusion (low signal on ADC, high signal on long b-value DWI) and mild hypervascularity on DCE. Even if these characteristics overlap with PCa, as for prostatitis the changes are generally not as pronounced nor as focal as in PCa, and the capsule may be focally retracted.
- *Fibrosis.* Generally, it occurs as a result of inflammation with subsequent scarring, and is characterized by band-like signal changes of low T2, no significant restriction of diffusion nor hypervascularity.
- *Cysts.* Have a typical appearance as a fluid-filled, well-demarcated structures of high T2 and low T1 signal. If they contain proteinaceous/haemorrhagic fluid, the signal can be variable (typically high T1 signal and variably reduced T2 signal). They usually don’t cause diagnostic dilemmas.
- *Haemorrhage.* It is more commonly encountered after biopsy and it is characterized by hyperintense signal on T1 and hypointense signal on T2 sequences, while chronic blood products may appear hypointense on all MRI

sequences. When diffuse, it may confound mpMRI assessment and mask small/low-grade foci of PCa; as a consequence, MRI should be performed after an adequate amount of time after biopsy procedures, usually at least 6 weeks. Blood in the seminal vesicles can be seen even in patients who did not undergo biopsy and may manifest itself with hematospermia.

Technical specifications

MRI studies are tailored to specific clinical questions, patients, time slots and available MRI equipment. Prostate MRI protocols should be limited to essential sequences for the diagnostic purpose and include small field-of-view (FOV) sequences centred on the prostate as well as at least one large FOV sequence on the pelvic region (i.e., up to the level of the aortic carrefour) to evaluate the pelvic lymph nodes.

The MRI scan quality is key for a correct evaluation. The supervising radiologist and the radiographer should join their efforts to maximise MRI quality, performing constant quality control. For example, it a particular sequence is affected by significant artefacts and is not diagnostic, the issue should be addressed, and the sequence repeated.

The recommended magnetic field strength to be used for prostate MRI is 1.5T or 3T. The theoretic advantage of 3T scanners is an increased signal-to-noise ratio (SNR), with positive effects on spatial/temporal resolution. However, magnetic field heterogeneity and susceptibility artefacts increase as well. As a consequence, a careful optimization of the acquisition parameters as well as the use of state of the art technology is required. 1.5T could be preferred in patients with devices that are either MR conditional at 1.5T only or when the location of the devices is likely to result in major artefacts (e.g., bilateral hip prosthesis).

Endorectal coils (ERCs) increase the SNR when combined with surface coils, with positive effects on spatial/temporal resolution and scan times, especially in larger patients. The major drawbacks of ERC are the increased procedural time and costs, potential gland deformation, introduction of artefacts (from air or the coil itself) and increased discomfort for the patients. When inflated with air, ERCs may cause distortion artefact that are more pronounced on DWI and on 3T scanners; to address this issue, inflation with liquids may be considered.

As recommended in the PI-RADS v2.1 guielines, ERC is required in older 1.5T scanner to obtain a satisfactory spatial/temporal resolution and to improve the diagnostic quality. Newer up-to-date 1.5T scanners using high density surface coils generally yield a satisfactory scan quality even in the absence of ERC.

Below are reported the technical parameters for mpMRI as suggested by the PI-RADS v2.1 recommendations (Turkbey *et al*, 2019).

T2W imaging. Should be obtained at least in the axial plane and in one of the orthogonal planes (sagittal/coronal). Planes can be pure or oblique (i.e., orientated along the axis of the prostate). Preferred technique is the 2D RARE (or FSE or TSE), without excessive echo train lengths to avoid blurring.

- Slice thickness: 3mm, no gap.
- FOV: 12-20 cm to (should include the prostate gland and seminal vesicles)
- In plane dimension: ≤0.7mm (phase) x ≤0.4mm (frequency)

3D axial acquisitions may be used, knowing that the soft tissue contrast may be reduced and the in plane resolution may be lower.

DWI imaging. DWI sequences should include high-b values images (>1400 sec/mm2; either acquired or synthetic) and ADC maps. To calculate ADC maps, two b-values under b1000 can be used with a monoexponential model.

Free-breathing spin echo EPI sequence combined with spectral fat saturation is recommended.

- Slice thickness: ≤4mm, no gap.
- TE: ≤90 msec; TR: ≥3000 msec
- FOV: 16-22 cm
- In plane dimension: ≤2.5mm phase and frequency

If two b-values only are acquired for ADC calculation it is recommended to use one low b-value (50-100 sec/mm2) and one intermediate b-value (800-1000sec/mm2). The highes b-value should be always ≤1,000 sec/mm2 to avoid diffusion kurtosis. The high b-value acquisition is mandatory and should preferably be obtained from a separate acquisition or calculated from the low and intermediate b-value images.

Dynamic Contrast-Enhanced (DCE) imaging. It is defined as the rapid T1W acquisition before, during and after IV administration gadolinium-based contrast agent (GBCA). PCa often demonstrate early enhancement compared to normal tissue. However, the kinetics of enhancement is heterogeneous (peak, washout, etc…). Early enhancement alone does not implicate the presence of csPCa and its absence does not exclude the possibility. To date, the added value of DCE above T2W + DWI is thought to be modest, and its role in in determination of PI-RADS score is secondary to T2W and DWI.

DCE is considered positive when focal, earlier (or contemporaneous) with that of normal prostatic tissue, usually corresponding to suspicious findings on the other

sequences. In the TZ, DCE has low specificity since BPH nodules may show early enhancement.

- Temporal resolution: <15 s
- Fat suppression and/or subtractions
- 3D T1W GRE generally preferred
- FOV: to include prostate and SV
- Slice thickness: 3mm, no gap
- TR/TE: <100msec/ <5msec
- In plane dimension: ≤2mm
- Total observation time: >2min
- GBCA dose: 0.1 mmol/kg at 2-3 cc/sec injection rate.

Assessment and reporting

Prostate MRI assessment and reporting should be standardized to increase the consistency across different centres and readers and to facilitate the communication with other specialists. Two scoring systems are commonly adopted in clinical practice: PI-RADS and Likert.

PI-RADS. In 2012, the European Society of Urogenital Radiology (ESUR) published guidelines for a correct and standardized prostate MRI acquisition / assessment / reporting, including a novel scoring system called as Prostate Imaging-Reporting and Data System version 1 (PI-RADS v1) (Barentsz *et al*, 2012). The main goal of PI-RADS is "to promote global standardization and diminish variation in the acquisition, interpretation, and reporting of prostate mpMRI examinations" (Turkbey *et al*, 2019). This standardization is not only critical for the assessment of the MRI findings, but also for communications, management, and research.

PI-RADS v1 was based on defined criteria for the assessment of T2, DWI and DCE, to assign a final score on a scale from 1 to 5 to each lesion (or region of the prostate), where 1 means very low and 5 very high probability of csPCa. PI-RADS v1 has been broadly validated but several limitations became apparent, leading to an update and improvement of the scoring system, the PI-RADS v2. The main update of the PI-RADS 2 was the introduction of the concept of "dominant sequence", that is the sequence that mostly determines the final score depending on the location of the lesion (PZ: dominant sequence DWWI; TZ: dominant sequence T2). Also, the role of DCE was recognised to be less relevant than previously, mainly to assess equivocal findings in the PZ. The role of PI-RADS v2 for the detection of csPCa was validated by numerous studies but, as expected, the system still showed some limitations mainly regarding the interreader variability, especially in the assessment of TZ. The

last (minor) update of the PI-RADS scoring system came in 2019, with the version PI-RADS v2.1, that mostly addressed those limitations in variability of the PI-RADS v2 lexicon, also bringing some updates and refinement of technical requirements.

Likert. Likert is a subjective and generic scale of likelihood that in prostate MRI is used to define the probability of harbouring significant disease, where 1 means very low probability and 5 very high probability. In Likert, the same lexicon and qualitative descriptions of PI-RADS are used for PCa assessment, but there are some significant differences. First, clinical information (e.g., PSA, PSAd, age, DRE) is used to define the final score along with the imaging features (Latifoltojar *et al*, 2019). Second, Likert does not adhere to the concept of dominant sequence as for PI-RADS (i.e., DWI for PZ and T2W for TZ), and any sequence can influence the final score more flexibly, especially in the TZ where the use of T2W as dominant sequence has been questioned. Third, the use of DCE is not limited to the upgrade of equivocal scores in the PZ and does not refer to focal lesions only. Finally, the distinction of 4 and 5 categories are not limited to dimensional (> 1.5 cm) criteria. In all, those difference make Likert a more flexible option for scoring prostate MRI especially in expert hands. On the other hand, one concern is that a greater freedom may bring greater interreader variability in scoring, but several studies have found comparable agreement of PI-RADS and Likert (Brizmohun Appayya *et al*, 2018).

In all, PI-RADS has gained more popularity and more acceptance worldwide both in clinical practice and in research. For this reason, a comprehensive description of PI-RADS scoring system will be provided.

PI-RADS v2.1 scoring system

(Turkbey *et al*, 2019) The main objective of prostate MRI is to detect, localize and characterize any abnormality that correspond to csPCa. The latest version is PI-RADS v2.1. PI-RADS uses a 5-point scale based on the likelihood of the presence of csPCa in a determined area of the prostate where:

- PIRADS 1 – Very low probability
- PIRADS 2 – Low probability
- PIRADS 3 – Intermediate probability: the presence of csPCa is equivocal
- PIRADS 4 – High probability
- PIRADS 5 – Very high probability

This classification is based on mpMRI findings only and should not take into account clinical information (e.g., PSA, DRE, family/clinical history, etc…). It is generally accepted that PI-RADS 1-2 don’t need biopsy while PI-RADS 4-5 lesions

should be biopsied, while for PI-RADS 3 lesions other factors could be taken into account. However, PI-RADS doesn't include recommendations for biopsy.

As introduced by PI-RADS v2, the final assessment relies on "dominant" sequences for each zone: DWI for the PZ, T2 for the TZ.

Measurement of Prostate Gland Volume. Any mpMRI report should include the volume of the prostate gland. It can be calculated using manual/automatic segmentation or manually using ellipsoid formula (AP x CC x LL diameters x 0.52). The PSAd can be then calculated as follows: PSA (ng/ml)/prostate volume (ml).

Mapping Lesions. PCa is tipically multifocal, but the largest tumour focus usually harbour the highest grade tumour, introducing the concept of "index lesion". According to PI-RADS v2.1, up to 4 lesions with 3-4-5 scores should be reported (ideally using the sector map) and the index lesion identified. The index lesion corresponds to that with the highest PI-RADS score; if more lesions have the same PI-RADS, the index lesion should be considered the bigger and/or the one with more aggressive features (i.e., ECE or SVI).

Measurement of lesions. MpMRI has several limitations in the assessment of lesion volume. First, lesion conspicuity (and volume) may vary across different sequences. Also, the ideal plane for measuring the lesions has not been determined. Finally, it has been shown that MRI consistently underestimates lesion volume compared to histology, especially the Gleason 3 component. However, PI-RADS guidelines contain recommendations to help standardizing lesion measurement. Any lesion should be measured at least using the larger dimension on axial images, and in any additional plane if the largest dimension is not in the axial plane. Alternatively, or in adjunction, lesion volume can be calculated using segmentation tools or with the ellipsoid formula (AP x CC x LL x 0.52). In the PZ, ADC should be preferred for measuring the lesions; in the TZ, T2 sequences are preferred.

Evaluation of the TZ. BPH is almost invariably present in men undergoing prostate MRI, altering the background for the assessment of TZ tumour. The margins of any suspicious nodule should be assessed in at least two planes on T2 sequences. Any finding should be scored that has different imaging characteristics from the nodular background. The T2W is the dominant sequence that determines the PI-RADS score in the TZ. Typical glandular (hyperintense) nodules are benign and should not be reported. Encapsulated (or mostly encapsulated nodules) with low T2 signal atypical

or stromal nodules (score of 2). Occasionally, atypical nodules may contain cancer and manifest with markedly restricted diffusion (and differing from other nodules). Consequently, they should be upgraded to PI-RADS 3 if a DWI score of ≥4. Stromal nodules frequently have moderately restricted diffusion.

Evaluation of the Central Zone. CZ tumours are uncommon and generally arise in the adjacent PZ or TZ. The normal CZ may have variable appearances also related to the presence of BPH that compresses or displaces the CZ. The normal CZ has bilaterally low signal intensity on T2 and ADC around the ejaculatory ducts from the base to the verumontanum. It may be symmetrically mildly hyperintense on DWI, without early enhancement or asymmetric signal on DWI. Early enhancement and/or asymmetry on T2-DWI-ADC may be related to the presence of PCa. Size asymmetry alone may be a normal variant especially in the presence of BPH. Rarely, CZ may be found as a discrete midline nodule above the verumontanum.

Anterior fibromuscular stroma (AFMS). Normal AFMS is a symmetric anterior structure with a crescentic shape characterized by low SI on T2W, ADC, and DWI (similar to that of obturator muscles), without early enhancement. Tumour in the AFMS may manifest with increased (and often asymmetric) SI on T2 and DWI with early enhancement, with or without evidence of a focal mass.

Caveats for overall assessment.

- Prostatitis may mimic tumour SI on all sequences. Morphology and SI may be used to assess the likelihood of malignancy, with indistinct/linear/diffuse changes less likely to be malignant.
- DWI is the dominant sequence in the PZ
- T2W is the dominant sequence in the TZ
- Detection and characterization of PCa is less reliable in the TZ compared to the PZ
- Bilateral symmetric signal changes often due to normal anatomy or benign changes

Multiparametric MRI assessment

T2 weighted imaging. T2W sequences are used to assess the prostate anatomy, detect any abnormality of the gland (benign or malignant) and for loco-regional staging (ECE, SVI, lymph node involvement).

Typical appearance of PCa in the PZ is a focal, round or ill-defined lesion with hypointense signal. However, significant overlap exists with benign entities such as prostatitis, atrophy, scars, haemorrhage, ectopic BPH with stromal component.

Typical appearance of PCa in the TZ is a focal, non-encapsulated, irregular or lenticular lesion with ill-defined borders ("erased charcoal" sign), eventually with aggressive features (i.e. invasion of adjacent structures). In the TZ, MRI is even less specific because the background signal is usually very hetherogeneaous and a significant overlap exists with benign entities, in particular BPH with stromal component.

Of note, benign changes could mimic or obscure csPCa both in the PZ and the TZ on T2W images. Tumours, especially bigger ones, may cross boundaries and involve both the PZ and the TZ.

The PI-RADS assessment on T2W is shown in Table 2.

Score	Peripheral Zone (PZ)
1	Uniform hyperintense signal intensity (normal)
2	Linear or wedge-shaped hypointensity or diffuse mild hypointensity, usually indistinct margin
3	Heterogeneous signal intensity or non-circumscribed, rounded, moderate hypointensity Includes others that do not qualify as 2, 4, or 5
4	Circumscribed, homogenous moderate hypointense focus/mass confined to prostate and <1.5 cm in greatest dimension
5	Same as 4 but ≥1.5cm in greatest dimension or definite extraprostatic extension/invasive behavior

Score	Transition Zone (TZ)
1	Normal appearing TZ (rare) or a round, completely encapsulated nodule. ("typical nodule")
2	A mostly encapsulated nodule OR a homogeneous circumscribed nodule without encapsulation. ("atypical nodule") OR a homogeneous mildly hypointense area between nodules
3	Heterogeneous signal intensity with obscured margins Includes others that do not qualify as 2, 4, or 5
4	Lenticular or non-circumscribed, homogeneous, moderately hypointense, and <1.5 cm in greatest dimension
5	Same as 4, but ≥1.5cm in greatest dimension or definite extraprostatic extension/invasive behavior

***Table 2**. PI-RADS T2WI assessment (Turkbey et al, 2019).*

DWI. Diffusion weighted sequences assess the extent of the free motion of water molecules within the tissues, and are central in the evaluation of prostate cancer. The assessment is made both on high b-value images and ADC maps. The ADC map is a visual display of the calculated values of water diffusion on each voxel (derived from DWI sequences), in μm2/sec.

Typically, csPCa is characterized by restricted water molecules diffusion that manifests as high signal on high b-value images and low signal on ADC map. The ADC values have been described to correlate with Gleason grade; however, absolute

values are known to be affected by technical parameters (b-values, exponential models, etc…) and can be inconsistent across vendors. As a consequence, the preferred method for ADC assessment is qualitative, even if a threshold of 750-900 μm2/sec can be used to distinguish cancer from benign pathology.

High b-value images are central in the qualitative evaluation of lesion diffusion, since lesion conspicuity can be increased compared to ADC maps especially in areas with lower SI (e.g. AFMS), in subcapsular lesions or in lesions located at the extreme apex/base of the gland.

PI-RADS assessment of DWI/ADC is shown in Table 3.

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 the 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 hypontense on ADC and markedly hyperintense on high b-value DWI; <1.5cm in greatest dimension
5	Same as 4 but ≥1.5cm in greatest dimension or definite extraprostatic extension/invasive behavior

Table 3. *PI-RADS DWI/ADC assessment (Turkbey et al, 2019).*

When assessing DWI/ADC, there are some caveats to be aware of:

- Findings on DWI/ADC should be correlated to the other sequences (T2W and DCE).
- Some benign entities may show low signal on ADC and T2W due to lack of overall signal (calcifications, blood products, fibrosis), but they usually show low signal on DWI sequences too.
- DWI/ADC can be limited in the assessment of BPH nodules that may show restricted diffusion, leading to equivocal findings. T2W images should be evaluated for the presence of capsule and other benign signs.
- Clearly encapsulated nodules in the PZ or CZ is likely an ectopic BPH nodule, and should be scored as PI-RADS 2.

Dynamic contrast enhanced imaging (DCE). PCa typically shows increased and earlier contrast enhancement (CE) compared to the normal glandular tissue, eventually with rapid washout. However, the kinetics of CE is variable and shows

significant overlap between benign/malignant processes. Also, the absence of CE alone is not reliable to exclude the presence of csPCa.

The accepted definition of positivity on DCE images is the presence of any area of focal, earlier CE compared to background enhancement that correspond to suspicious areas on T2W and DWI. It usually appear within 10 second of the appearance of CM in the femoral arteries.

However, to date the added value of DCE is still debated (over T2W and DWI), and its value in the final assessment of PI-RADS scores is still limited. However, DCE may sometime assist in the detection of PCa that is occult on T2W and DWI, and in clinical practice is commonly used as a "safety net" especially when the quality of the other sequences (especially DWI) is degraded by artefacts.

PI-RADS assessment of DCE is shown in Table 4.

Score	Peripheral Zone (PZ) or Transition Zone (TZ)
(–)	no early or contemporaneous enhancement; or diffuse multifocal enhancement NOT corresponding to a focal finding on T2W and/or DWI or focal enhancement corresponding to a lesion demonstrating features of BPH on T2WI (including features of extruded BPH in the PZ)
(+)	focal, and; earlier than or contemporaneously with enhancement of adjacent normal prostatic tissues, and; corresponds to suspicious finding on T2W and/or DWI

Table 4. *PI-RADS DCE assessment (Turkbey et al, 2019).*

PI-RADS and biparametric MRI (bpMRI). Evidence on the limited utility of DCE in the detection of PCa raised the possibility of the use of simplified MRI protocols without the use of contrast media (and DCE), commonly referred to as bpMRI.

There are some potential benefits of bpMRI: the elimination of potential CM side effects, reduced procedural times, reduced costs, higher accessibility and higher acceptability. Of note, its role in the determination of final PI-RADS score is formally limited to equivocal findings in the PZ.

However, some concerns exist over bpMRI. Some studies have shown that sensitivity may be improved by DCE, especially when the quality of DWI is impaired by artefacts; furthermore, DCE can be used as a safety net, increasing confidence in MRI interpretation and scoring; finally, it has been reported that bpMRI can be associated with a higher rate of equivocal findings, reducing the benefits of MRI as a triage test for biopsy and decreasing the confidence in the MRI-directed pathway for PCa diagnosis.

In all, the PI-RADS guidelines highlight the need for large, prospective trials before bpMRI is implemented in clinical practice, and recommend the use of mpMRI as the standard of care especially when the focus is not to miss any csPCa. Also, mpMRI may be preferred in all cases before focal/radical treatment, for a better follow-up evaluation especially when the morphology of the prostate gland is likely to be altered. Final PI-RADS score assessment is shown in Table 5-6.

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

Table 5. *PZ PI-RADS assessment (Turkbey et al, 2019).*

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

Table 6. *TZ PI-RADS assessment (Turkbey et al, 2019).*

The role of MRI in clinical diagnosis of PCa

The role of mpMRI of the prostate in the diagnostic workup of PCa has evolved dramatically in the past decade.

PSA dosage, while debated, still remains the gold standard for guiding PCa screening and early detection (Mottet *et al*, 2020). Historically, men with clinically suspected PCa (i.e., elevated PSA, positive DRE, familiar history) were candidates to transrectal ultrasound guided (TRUS) systematic biopsies (SBx). TRUS-SBx consist in random systematic sampling of the whole prostate gland, obtaining at least 8 biopsy cores for small prostates and up to 12 cores for large prostates (Mottet *et al*, 2020). This approach had several limitations. First, PSA is not specific to PCa, leading to biopsy in many cases of benign pathology (e.g., BPH, inflammation) (Grossman *et al*, 2018). Second, due to sampling errors, up to 30% of PCa may be missed especially in the TZ and prostatic apex (Guichard *et al*, 2007). Third, a significant number of indolent tumours may be identified that would not lead to significantly

increased morbidity/mortality but cause unnecessary patient distress and may lead to overtreatment. Finally, TRUS-SBx are associated with potential complications, especially infections, with a rate of serious adverse events (needing hospitalization) of >1%. As a consequence, additional tests were advocated to minimize the number of men undergoing unnecessary biopsies while increasing the detection of csPCa.

After the introduction of mpMRI of the prostate, there has been a growing body of evidence that mpMRI had high accuracy in the detection of PCa. In 2017, the results of the PROMIS study were published (Ahmed *et al*, 2017). PROMIS was a prospective, multicentre, paired-cohort study to assess diagnostic accuracy of mpMRI and TRUS-Sbx using template transperineal (TTPM) mapping biopsies as the reference test. MpMRI proved to be more sensitive than SBx. Using mpMRI before biopsy would have spared biopsy to approximately one third of men, with 5% fewer indolent cancers detected; at the same time, targeted biopsies would have allowed to detect 18% more csPCa. The PROMIS study thus showed that mpMRI, if used as a triage test for biopsy, could significantly reduce the number of unnecessary biopsies and the number of non-csPCa tumours detected, while improving the detection of csPCa.

Following the success of PROMIS, the PRECISION study was published in 2018 (Kasivisvanathan *et al*, 2018). The PRECISION was a multicentre, randomized, noninferiority trial, in which 500 men with clinically suspected prostate cancer were randomized in two groups: the MRI-targeted biopsy group (MRI-TBx) where men underwent MRI +/- targeted biopsy, and the SBx group where men underwent standard 10-12 cores TRUS-Bx. In the MRI-TBx group, 28% had a negative MRI and consequently did not undergo biopsy. csPCa was identified in 38% of men of MRI-TBx group and in 26% of men in the SBx group; 13% fewer indolent PCa were identified in MRI-TBx group vs SBx group. In summary, the PRECISION study demonstrated that mpMRI performed before biopsy +/- MRI-TBx clearly outperforms the standard diagnostic pathway based on SBx alone.

Another prospective multicenter, paired diagnostic study (MRI-FIRST) was conducted in men who underwent mpMRI +/- TBx and SBx in all men regardless the MRI results. In MRI-FIRST, no difference between SBx and TBx was observed for the detection of ISUP $\geq$2 disease, but the detection was improved by combining both techniques, showing that mpMRI before biopsy improves the detection of csPCa but does not seem to avoid the need for systematic biopsy (Rouvière *et al*, 2019).

More recently, a Cochrane Systematic Review and Meta-analysis was published on the diagnostic accuracy of MRI and SBx (Drost *et al*, 2019), showing a pooled sensitivity of 91% and a pooled specificity 37% of mpMRI, with overall diagnostic

advantage of MRI over SBx; however, the quality of the studies included and their heterogeneity limit the interpretation of the results.

The above-mentioned results apply to biopsy naïve men. In a repeated biopsy setting, where previous negative SBx have been performed and have given negative results, the advantage of performing MRI before re-biopsy is even more pronounced. In the PICTURE study (Simmons *et al*, 2017), where men requiring repeat biopsy underwent mpMRI and TTPM-biopsies, men with negative mpMRI could have safely avoid re-biopsy (NPV 91%). Accordingly, the Cochrane meta-analysis of Drost et al. (Drost *et al*, 2019) showed that MRI-TBx outperforms SBx for csPCa detection in the repeat-biopsy setting, while the difference in biopsy-naïve men is less pronounced.

In all, evidence on MRI performance led to a paradigm shift in the diagnostic workup of PCa. The EAU guidelines on prostate cancer (Mottet *et al*, 2020) now recommend that mpMRI is performed prior to biopsy in men with a clinical suspicion of PCa, both in biopsy-naïve men and in a repeated biopsy setting (Level of evidence: 1a). Biopsies should be performed in the presence of suspicious lesion at MRI (PI-RADS > 2), both targeted and systematic. In the presence of negative mpMRI, biopsy could be omitted based on a shared decision making with the patient.

Prostate MRI for population-based screening

There has been a significant debate about the role of PSA for population-based PCa screening in the last decade (Eldred-Evans *et al*, 2020). Specifically, despite an observed reduction of mortality from PCa, PSA screening leads also to significant overdiagnosis and overtreatment. After the development of ultra-fast, non-contrast MRI acquisition protocols (i.e., biparametric MRI), interest has grown over MRI as a potential tool for primary screening, analogous to mammography for breast cancer and low-dose CT for lung cancer. Recently, two prospective studies in the UK have investigated the feasibility of MRI for population-based screening (Eldred-Evans *et al*, 2021; Marsden *et al*, 2021b), yielding encouraging results. However, there are still several limitations to be addressed before MRI can be considered for screening. First, despite biparametric MRI can be as accurate as mpMRI for PCa detection (Zawaideh *et al*, 2020), evidence on the performance of ultra-fast bpMRI protocols is still limited (van der Leest *et al*, 2019b). Also, the optimal MRI protocol is still to be determined (e.g., abbreviated biparametric MRI), as well the optimal strategy for interpretation (e.g., single vs double-reads) and the optimal MRI positivity cut-off. Further investigations are needed before the implementation of MRI in national screening programmes.

Limitations of MRI

While MRI has revolutionised the diagnosis of PCa, there are still several limitations that need to be addressed.

Accessibility and costs. Multiparametric MRI is associated with long acquisition and procedural times. Scan times are variable depending on institutional protocols, but are generally around 30 minutes (Brembilla *et al*, 2022). Procedural times include the need for pre-scan paperwork, cannulation, endorectal coil placement, post-scan observation. The need for contrast media also requires additional dedicated personnel to be present at the time of MRI scan. Long scan times result in reduced number of exams that can be performed per day/machine. This, together with overall limited availability of dedicated scanner and personnel (radiologists, radiographers, nurses), leads to limited accessibility to MRI in routine practice and long waiting times. The need for long scan/procedural times, together with the required personnel, come with elevated costs of MRI (van der Leest *et al*, 2019b).

Contrast media. Multiparametric MRI requires the use of gadolinium-based contrast media (GBCM). The use of GBCM can be associated with adverse events, especially allergic reactions, even if rare. Concerns have also been raised regarding the deposition of Gadolinium in the brain, especially after multiple administrations. Also, the use of GBCM increases costs, procedural times and requires additional dedicated personnel. However, the utility of DCE in the detection of csPCa has been challenged by recent reports, proposing bpMRI as a valuable alternative to avoid the need of contrast media (Bass *et al*, 2020).

False positive results. The specificity of MRI is consistently low across different studies, with a recent metanalysis showing a pooled specificity of 37% (Drost *et al*, 2019). In a typical diagnostic setting, positive predictive value (PPV) is low, with most papers reporting an overall value under 50% (Westphalen *et al*, 2020), decreasing with lower PI-RADS scores (Mazzone *et al*, 2021). This means that approximately 1 in 2 men undergoes unnecessary targeted biopsies. Several strategies have been explored to reduce the number of false positive findings. Among the others, the use of PSA density improves the detection rate of PCa reducing the number of unnecessary biopsies especially for equivocal scores (PI-RADS 3). However, an improvement of MRI technique and, possibly, the implementation of new sequences should be explored to address the low specificity of MRI.

False negative results. One of the main strength of mpMRI is its high sensitivity (generally >90%) (Drost *et al*, 2019). In a typical diagnostic setting, also NPV is high and around 90% (Ahmed *et al*, 2017), meaning that 9 negative MRI scans out of 10 do not harbour significant cancer. Despite its high sensitivity, prostate MRI is still

prone to missing significant PCa, with a wide variation in proportions of overlooked PCa described in the literature (ranging approximately from 10 to 50%), depending on study design and standard of reference. Even if the majority of overlooked PCa are low volume and low grade (Norris *et al*, 2020), there is still room for improvement of prostate MRI to detect less conspicuous PCa, ideally using novel sequences or techniques that might help to better characterise less conspicuous tumours on mpMRI.

Interobserver variability. PI-RADS has been introduced to standardize the acquisition and interpretation of prostate MRI. Still, the literature reports modest levels of reproducibility of prostate MRI using PI-RADS scores (Rosenkrantz *et al*, 2016; Greer *et al*, 2017). This observation at least partially crashes with observations from a routine clinical practice, where positive (PI-RADS 4-5) and negative (PI-RADS 1-2) cases are frequently concordant among radiologists (especially experienced ones). In fact, most of the studies available in the literature showing low reproducibility of MRI have significant flaws in study design that may limit the interpretation of the results. As a consequence, a more accurate evaluation of the real interobserver reproducibility of MRI is advocated to have a realistic estimation of MRI reproducibility. This is of paramount importance to drive future developments of scoring systems and to compare the reliability of new MRI sequences or imaging techniques with the standard prostate MRI protocol.

MRI quality. Achieving an MRI scan of good quality is of paramount importance to reliably rule-in/rule-out PCa (Giganti *et al*, 2020). It has been described how scan quality affects MRI scores (Karanasios *et al*, 2022). Specifically, diffusion weighted imaging (the key functional imaging sequence of mpMRI) often suffers from image distortion/degradation (moderate/severe in approximately 40% of studies) due to artefact caused by air within the rectum causing difficulty in scan interpretation (Caglic *et al*, 2017). Thus, there is a need for improved image quality and quality control of MRI, via standardization of scan protocols and potentially via the investigation of novel MRI techniques that are more resistant to the typical image artefacts.

Aim of the work

The present project is aimed to address some of the main limitations of prostate MRI, specifically to explore the potential for simplification of the standard MRI protocols in clinical routine and its application to novel settings (i.e., screening), and to validate novel MRI techniques that could allow a better detection and characterization of prostate cancer.

In details, the main objectives of the study are:

- To determine the interobserver agreement of prostate MRI interpretation and scoring among readers of various experience in a real-world setting. This will serve as a reference for the evaluation of the variability of alternative MRI protocols or novel MRI techniques compared to the standard of care (i.e., multiparametric MRI).
- To investigate the feasibility of prostate MRI protocol simplification, assessing the diagnostic performance and interobserver variability of fast/ultra-fast MRI protocols (i.e., biparametric MRI and abbreviated biparametric MRI) for PCa detection and screening. Also, the best scoring approach for a screening setting will be investigated (e.g., single vs double-readings, optimal MRI positivity cut-off).
- To validate a novel MRI technique called Luminal Index MRI (LI-MRI), exploring its feasibility, reproducibility and accuracy in men with clinically suspicious prostate cancer (LI-MRI), and to explore its applicability in a real-world clinical setting.

Interreader Agreement of Prostate Imaging Reporting and Data System Version 2.1

(The content of this chapter has been published by Brembilla et al. 2020, DOI: https://doi.org/10.1007/s00330-019-06654-2).

In men with clinically suspected prostate cancer (PCa), multiparametric magnetic resonance imaging (mpMRI) reduces the number of unnecessary biopsies and improves the detection of clinically significant PCa (csPCa), while decreasing the detection of clinically insignificant PCa (non-csPCa) (Kasivisvanathan *et al*, 2018; Ahmed *et al*, 2017). Recently, international urological guidelines recommended mpMRI before prostate biopsy both in biopsy-naïve patients and in patients with prior negative biopsy (Mottet *et al*, 2019; Drost *et al*, 2019). Given the widespread adoption of prostate MRI in clinical practice, Prostate Imaging Reporting and Data System (PI-RADS) has been introduced to standardize interpretation and reporting of MRI examinations (Barentsz *et al*, 2012; Weinreb *et al*, 2016). Specifically, PI-RADS v2.1 has addressed limitations in interreader agreement of the previous versions (Turkbey *et al*, 2019). Prior reports revealed only moderate reproducibility of PI-RADS v2 (Rosenkrantz *et al*, 2016), with poor to moderate agreement in lesion detection (Schimmöller *et al*, 2013; Muller *et al*, 2015; Sonn *et al*, 2018; Pickersgill *et al*, 2018; Smith *et al*, 2019; Girometti *et al*, 2019). However, these studies had several limitations in methodology (e.g.: interpretation of screen captures) and in patients selection (only biopsy or radical prostatectomy cohorts) that may prevent the generalizability of their findings to a real-life clinical setting. Moreover, the statistical approach commonly used (*k* coefficient) is known to be exposed to severe paradoxes in determined circumstances, (Feinstein & Cicchetti, 1990; Cicchetti & Feinstein, 1990) potentially underestimating the true extent of the agreement (Shih *et al*, 2018).

To overcome these issues, we evaluated interreader agreement of prostate mpMRI using PI-RADS v2.1 in a cohort of patients referred for prostate MRI at our institution for clinically suspected PCa, reproducing the typical clinical workflow. In this setting, we investigated the reproducibility of multiple readers with different expertise in lesion detection, and we evaluated which clinical and radiological features may influence variability. We hypothesized that, using proper statistical analyses in a non-selected cohort of patients, observed agreement may be higher than previously reported.

Materials & methods

Study Population

This retrospective study was approved by our Institutional Review Board and written informed consent was obtained from all patients. Our prospectically acquired local database was used to identify 219 consecutive biopsy-naïve or previous negative biopsies men who underwent prostate mpMRI at our institution (San Raffaele Hospital, Milan) for a clinically suspected PCa between May and September 2017. All MRI examinations were formerly interpreted and scored using PI-RADS v2 by one of six dedicated radiologists at our institute. Pre-MRI clinical information (PSA values, digital rectal examination, familiar history, previous local treatments) were collected for all patients. We subsequently excluded patients who had previous transurethral resection of the prostate (n=4), incomplete mpMRI protocol (n=1), low image quality or severe image artifacts (n=3) and missing clinical information (n=11). The final study cohort consisted of 200 men. Of those 70.5% (n=141) and 29.5% (n=59) were biopsy naïve and prior negative biopsies, respectively.

MRI Acquisition

MRI images were acquired on a 1.5-T scanner (Achieva and Achieva dStream, Philips Medical Systems) with surface and endorectal coil (Prostate eCoilTM, Medrad®); acquisition protocols were in line with PI-RADS v2 standards (Weinreb *et al*, 2016). Gastrointestinal peristalsis was suppressed by intramuscular administration of 20 mg of scopolamine-butylbromide (Buscopan, Boehringer) immediately before MR scanning. The imaging protocol consisted of multiplanar turbo spin-echo T2-weighted images, echo-planar DWI with b values of 50, 800 and 1600 s/mm2 (ADC maps were automatically elaborated on a pixel-by-pixel basis using b values of 50 and 800 s/mm2), 3-D fast field-echo dynamic contrast- enhanced (DCE) MRI and delayed axial turbo spin echo T1- weighted images with fat suppression. For DCE-MRI, an IV bolus of 0.1 mmol/kg of gadobutrol (Gadovist, Bayer Schering Pharma) at a flow rate of 4 ml/s was injected. For patients who had previously undergone prostatic biopsies, mpMRI scans were performed at least after 4 weeks from biopsies, and pre-contrast T1-weighted images were performed to rule out post-biopsy hemorrhagic artefacts.

MRI Interpretation

For interreader agreement analyses, all examinations were retrospectively reviewed, interpreted and scored according to PI-RADS v2.1 (Turkbey *et al*, 2019) by seven radiologists from a single tertiary care referral center. Four were dedicated

radiologist with specific clinical and research interest in prostate MR imaging and 4 to 8 years of experience in prostate MRI (referred to as "dedicated" readers). Three were abdominal radiologists who underwent a specific training but who were not specifically dedicated to prostate MRI in clinical routine (approximately <10 prostate MRI examination per month), and had <2 years of experience in the field (referred to as "non-dedicated" readers).

Study Design

To replicate as much as possible the typical prostate MRI interpretation workflow, the 7 radiologists had full access to anonymized MRI examinations on a PACS workstation (Figure 1). For each patient, readers were provided with all pre-MRI clinical information (age, PSA values, DRE, family history and pre-MRI biopsy status), while they were blinded to original MRI report and post-MRI information (e.g. biopsy results). After image interpretation, findings were reported on a standardized form (Appendix) that included: 1) presence/absence of equivocal or suspicious lesions (PI-RADS score ≥3); 2) PI-RADS v2.1 score for each lesion; 3) lesion localization and diameters; 4) a score for peripheral zone (PZ) signal intensity (SI) homogeneity, as proposed by Hötker et al. (Hötker *et al*, 2019) (a scale from 1 to 5 indicating the grade of homogeneity in the PZ, where 1 means markedly inhomogeneous SI and 5 indicates a highly homogeneous SI); 5) a subjective score on interpretative difficulty of the MRI images (scale from 1 to 3, where: 1: easy; 2: intermediate; 3: difficult). Readers were asked to provide screenshots of each PI-RADS score ≥3 lesion on T2W images, that were used to determine lesion-specific agreement. After all readers completed the reviewing process, a consensus revision was made for all cases to determine the radiological standard of reference.

Biopsy and Histopathology

The decision to perform or to avoid biopsy was made at the time of MRI examination and was based on the original report. All patients with at least one PI-RADS score ≥3 lesion at original MRI report (n=110, 55%) underwent targeted biopsy with fusion or cognitive approach, as previously described (Stabile *et al*, 2018). Each patient was also concomitantly submitted to a standard 12-core random systematic biopsy (TRUS-Bx) (Dell'Oglio *et al*, 2019). Patients with negative MRI (maximum PI-RADS score <3) underwent TRUS-Bx if deemed necessary by the treating physician (n=22, 11%). The remaining patients with negative MRI did not undergo prostate biopsy (n=68, 34%), neither immediately after MRI nor during routine follow-up. All prostate biopsy specimens were analyzed by dedicated uro-pathologists. Clinically

significant PCa (csPCa) was considered as presence of any Gleason ≥3+4 (ISUP grade ≥2) at biopsy.

Biopsy results were then used to perform analyses on subgroup of patients harboring (or not) csPCa. It has to be noted that, since targeted biopsies were performed based on the original MRI report, their results could not correspond to the lesions identified after review using PI-RADS v2.1. Thus, all analyses based on biopsy should be considered on a per-patient level rather than a per-lesion level (i.e. agreement in men harboring or not csPCa).

Statistical Analysis

Medians and interquartile ranges, as well as frequencies and proportions were reported for continuous and categorical variables, respectively. Interreader agreement for multiple readers was evaluated using Conger's generalized kappa coefficient (Conger, 1980), contextually reporting raw percent agreement (PA).

Similarly to the others kappa coefficients, Conger's kappa is exposed to the same well-known paradoxes (Feinstein & Cicchetti, 1990; Cicchetti & Feinstein, 1990); in particular, k values can be unexpectedly low even in presence of high agreement, depending on marginal frequencies. To overcome this issue, we additionally calculated agreement with alternative methods. First, Agreement Coefficient 1 (AC1) was computed (Gwet, 2008): based on the assumption that chance agreement is likely to affect only a portion of the observations, and not relying on assumed independence between observations of the readers, AC1 is less prone to paradoxes than k coefficients. Second, we measured interreader agreement for presence or absence of lesions using indexes of specific positive and negative agreement (P_{pos} and P_{neg}) (Cicchetti & Feinstein, 1990). The mathematical calculation of P_{pos} is identical to the Index of Specific Agreement (ISA) proposed by Shih et al. (Shih *et al*, 2018), and represents the proportion of specific agreement relative to positive scores. Similarly, we calculated reader agreement on negative scores (negative MRI) by means of proportion of negative agreement (P_{neg}). Computing P_{pos} and P_{neg} allows to assess eventual differences in the agreement among readers on positive or negative cases. Even though these indexes are not chance-corrected (as k and AC1 coefficients), they are particularly suitable in this setting since the probability that more readers detect the same lesion in the same location by chance is negligible (Shih *et al*, 2018).

Levels of agreement were defined using the conventional classification of Landis & Koch (Landis & Koch, 1977): slight (0–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80) and excellent (0.81–1). Even though it was originally

proposed for k statistics, to simplify the comparison between different coefficients we extended this categorization also to PA, AC1 and indexes of specific agreement.

Statistical tests were performed using AgreeStat 2015.6 and RStudio graphical interface v.0.98 for R software environment v.3.0.2 (R Foundation, Vienna, Austria).

Subgroup Analysis

To evaluate specific features that may influence variability, agreement analyses in index lesion detection (using a positivity threshold of PI-RADS score ≥3) were performed in subgroups of patients defined upon clinical, bioptic or radiological parameters.

PSA values were considered relative to prostate volume and expressed as PSA density (PSAD); PSAD threshold was set at 0.15 ng/ml/cc (Rais-Bahrami *et al*, 2015). Pre-MRI clinical risk was assessed with ESPRC risk calculator 6 (Roobol *et al*, 2017); patients were then divided in quartiles according to calculated risk, and divided in low risk (≤25th percentile), intermediate risk (25-50th percentile) and high risk (≥75th percentile).

With regards to post-MRI biopsy results, we divided patients in three groups: patients harboring csPCa (ISUP group ≥2), patients harboring non-csPCa (ISUP group 1) and patients with negative biopsies (no PCa group). We included patients without post-MRI biopsy in no PCa group given the low risk of harboring PCa (Ahmed *et al*, 2017), and considering acceptable this approximation as diagnostic performance analyses were not a purpose of our study. Biopsy results were also used to assess the concordance in true positive (TP) Vs false positive (FP) findings at MRI. Then, in men who underwent biopsy, we compared the prevalence of csPCa in concordant-positive cases Vs discordant cases, using all possible pairwise combination of dedicated and non-dedicated readers.

With regards to MRI parameters, the radiological gold standard was used to define the presence or absence of index lesion in PZ and TZ and to determine the homogeneity score of the PZ (SI classified as "homogeneous" for homogeneity scores ≥3). MRI interpretation was classified as "difficult" when at least one dedicated reader gave a subjective interpretative difficulty score of 3/3, while it was considered "easy" when none of the readers gave a subjective interpretative difficulty score of 3/3. Multifocality was defined as the presence of more than one PI-RADS score ≥3 lesion at MRI.

Results

Clinical, radiological and pathological characteristics of the study cohort are summarized in Table 7.

Parameters	
No. of Patients	200
Age (y)	65 (58-70)
PSA (ng/mL)	6.0 (4.1-8.4)
Prostate volume (mL)	58.9 (42.4-79.2)
PSA density, PSAD (ng/mL/cc)	0.10 (0.07-0.16)
Pre-MRI biopsy status	
Biopsy naïve	67% (135/200)
Prior negative biopsy	33% (65/200)
Clinical Stage	
cT1	88% (176/200)
cT2,3	12% (24/200)
Positive familiar history	9% (17/200)
Original MRI report	
Negative MRI	45% (90/200)
Positive MRI	55% (110/200)
Post-MRI biopsy (positive MRI)	110
No PCa	47% (52/110)
Any PCa	53% (58/110)
csPCa	38% (42/110)
Post-MRI biopsy (negative MRI)	22
No PCa	86% (19/22)
Any PCa	14% (3/22)
csPCa	5% (1/22)
No biopsy (negative MRI)	68

Table 7. *Baseline characteristics of the study cohort (n=200). Values are reported as frequencies, medians (interquartile range in parentheses) or percentages (proportions in parentheses). Positive MRI: at least one PI-RADS ≥3 lesion; negative MRI: absence of PI-RADS ≥3 lesion. csPCa: ISUP group ≥2.*

Overall Agreement

Table 8 shows the agreement in assessing lesions at mpMRI with PI-RADSv2.1 based on kappa coefficient, AC1 and PA.

Overall, dedicated readers showed higher concordance than non-dedicated readers. Using a cut-off of PI-RADS score ≥3 for index lesion detection, agreement was moderate among all readers (k=0.591, 95%CI: 0.529-0.653) and non-dedicated readers (k=0.562, 95%CI: 0.481-0.643), while was substantial for dedicated readers (k=0.621, 95%CI: 0.548-0.694). Using a cut-off of PI-RADS >3, agreement was substantial for all readers, with highest scores among dedicated readers (k=0.779, 95%CI: 0.711-0.846).

Considering AC1 values, agreement was substantial in all groups of readers using a cut-off of PI-RADS score ≥3, while it was excellent with a cut-off of PI-RADS 4+5, with highest scores among dedicated readers (AC1=0. 876, 95%CI: 0.836-0.916).

Feature	All	Dedicated	Non-dedicated
All patients			
PI-RADS ≥ 3			
k	0.591 (0.529-0.653)	0.621 (0.548-0.694)	0.562 (0.481-0.643)
AC1	0.738 (0.695-0.782)	0.762 (0.713-0.811)	0.712 (0.652-0.771)
PA (%)	78.4 (74.9-81.9)	80.3 (76.3-84.3)	76.3 (71.5-81.1)
PI-RADS 4-5			
k	0.699 (0.634-0.764)	0.779 (0.711-0.846)	0.671 (0.589-0.753)
AC1	0.841 (0.806-0.878)	0.876 (0.836-0.916)	0.821 (0.772-0.870)
PA (%)	86.5 (83.5-89.6)	90.3 (87.3-93.4)	84.8 (80.8-88.9)
Clinical parameters			
Prev. Neg. Bx			
k	0.562 (0.435-0.690)	0.601 (0.430-0.771)	0.534 (0.388-0.679)
AC1	0.763 (0.687-0.839)	0.779 (0.675-0.884)	0.726 (0.628-0.824)
PA (%)	79.9 (73.7-86.1)	81.4 (72.8-89.9)	78.8 (71.7-85.9)
Biopsy Naive			
k	0.587 (0.513-0.661)	0.639 (0.552-0.727)	0.533 (0.438-0.629)
AC1	0.730 (0.677-0.783)	0.769 (0.709-0.829)	0.685 (0.612-0.758)
PA	77.8 (73.5-82.1)	80.9 (76.1-85.8)	74.2 (68.4-80.1)
PSAD >0.15			
k	0.671 (0.545-0.797)	0.734 (0.605-0.864)	0.595 (0.433-0.758)
AC1	0.797 (0.715-0.880)	0.822 (0.732-0.913)	0.743 (0.630-0.857)
PA	83.2 (76.4-90.0)	86.7 (80.0-93.3)	78.8 (69.6-88.0)
PSAD <0.15			
k	0.544 (0.473-0.616)	0.585 (0.490-0.680)	0.536 (0.440-0.631)
AC1	0.716 (0.664-0.767)	0.740 (0.674-0.806)	0.699 (0.627-0.771)
PA	76.5 (72.3-80.6)	78.5 (73.1-83.8)	75.2 (69.5-81.0)
Low Risk			
k	0.532 (0.493-0.662)	0.546 (0.377-0.714)	0.551 (0.364-0.739)
AC1	0.738 (0.649-0.826)	0.753 (0.649-0.857)	0.738 (0.602-0.873)
PA	77.9 (70.9-84.9)	79.1 (70.6-87.6)	78.0 (67.2-88.9)
Intermediate Risk			
k	0.510 (0.422-0.599)	0.536 (0.431-0.641)	0.483 (0-367-0.559)
AC1	0.681 (0.618-0.745)	0.703 (0.631-0.775)	0.655 (0.569-0.742)
PA	73.8 (68.7-78.9)	75.5 (69.6-81.4)	71.7 (64.9-78.7)
High Risk			
k	0.754 (0.640-0.869)	0.807 (0.682-0.932)	0.683 (0.536-0.831)
AC1	0.844 (0.766-0.921)	0.879 (0.797-0.961)	0.794 (0.689-0.898)
PA	87.1 (80.8-93.4)	90.0 (83.2-96.8)	83.0 (74.6-91.5)
MRI parameters			
PZ			
k	0.248 (0.141-0.354)	0.237 (0.093-0.380)	0.289 (0.137-0.441)

AC1	0.694 (0.615-0.774)	0.694 (0.592-0.796)	0.689 (0.589-0.789)
PA	73.1 (66.8-79.4)	74.4 (67.0-82.0)	72.8 (64.8-80.9)
TZ			
k	0.210 (0.058-0.362)	0.175 (0.033-0.383)	0.252 (0.052-0.557)
AC1	0.633 (0.461-0.805)	0.728 (0.549-0.906)	0.535 (0.294-0.776)
PA	68.2 (54.8-81.5)	75.4 (60.7-90.2)	61.4 (42.6-80.2)
Homogeneous SI			
k	0.766 (0.675-0.858)	0.804 (0.705-0.904)	0.677 (0.549-0.806)
AC1	0.839 (0.773-0.905)	0.807 (0.708-0.906)	0.767 (0.663-0.871)
PA	88.0 (83.2-92.9)	90.3 (85.3-95.2)	82.9 (75.4-90.3)
Inhomogeneous SI			
k	0.496 (0.420-0.573)	0.523 (0.430-0.617)	0.499 (0.396-0.602)
AC1	0.671 (0.615-0.728)	0.693 (0.626-0.760)	0.666 (0.588-0.744)
PA	73.0 (68.5-77.6)	74.7 (69.4-80.1)	72.7 (66.4-78.9)
Difficult			
k	0.185 (0.072-0.297)	0.191 (0.032-0.350)	0.138 (0.008-0.297)
AC1	0.460 (0.355-0.565)	0.436 (0.293-0.578)	0.387 (0.227-0.548)
PA	55.8 (47.8-63.7)	58.1 (48.5-67.8)	50.4 (38.4-62.4)
Easy			
k	0.685 (0.619-0.751)	0.718 (0.641-0.795)	0.665 (0.578-0.753)
AC1	0.811 (0.770-0.853)	0.833 (0.785-0.881)	0.796 (0.738-0.854)
PA	84.3 (80.9-87.7)	86.1 (82.1-90.0)	83.0 (78.3-87.8)
Multifocality			
k	0.508 (0.356-0-660)	0.475 (0.301-0.649)	0.432 (0.241-0.623)
AC1	0.721 (0.603-0.838)	0.718 (0.583-0.852)	0.644 (0.482-0.806)
PA	82.1 (76.1-88.1)	81.6 (74.5-88.7)	77.8 (69.4-86.2)
Post-MRI biopsy			
ISUP group ≥2			
k	0.370 (0.108-0.632)	0.394 (0.058-0.730)	0.443 (0.174-0.712)
AC1	0.859 (0.782-0.936)	0.888 (0.809-0.968)	0.843 (0.740-0.945)
PA	86.9 (80.1-93.7)	89.8 (82.9-96.6)	85.6(76.8-94.5)
ISUP group 1			
k	0.184 (0.071-0.439)	0.207 (0.061-0.475)	0.115 (0.024-0.425)
AC1	0.522 (0.344-0.701)	0.501 (0.271-0.731)	0.458 (0.149-0.768)
PA	59.9 (46.1-73.8)	61.8 (46.1-77.4)	54.9 (31.5-78.3)
No PCa			
k	0.478 (0.391-0.565)	0.501 (0.395-0.608)	0.456 (0.345-0.566)
AC1	0.744 (0.693-0.795)	0.764 (0.705-0.823)	0.693 (0.615-0.772)
PA	78.0 (73.9-82.1)	79.6 (74.7-84.5)	76.0 (70.4-81.7)
True Positive			
k	0.229 (0.079-0.378)	0.209 (0.021-0.439)	0.352 (0.085-0.619)
AC1	0.858 (0.780-0.936)	0.891 (0.814-0.967)	0.841 (0.737-0.944)
PA	86.6 (79.7-93.5)	89.5 (82.5-96.5)	85.3 (76.2-94.3)
False Positive			
k	0.144 (0.067-0.221)	0.198 (0.035-0.361)	0.115 (0.019-0.213)
AC1	0.528 (0.436-0.619)	0.533 (0.400-0.666)	0.529 (0.419-0.638)
PA	60.1 (53.2-67.0)	60.8 (50.5-71.1)	59.9 (51.6-68.3)

Table 8. *Agreement on Index Lesion detection of Prostate Imaging Reporting and Data System version 2.1. Values in parenthesis are 95% confidence intervals. PA: proportion of agreement; Prev. Neg. Bx: previous negative biopsy; PSAD: PSA density (expressed in ng/ml/cc); PZ: peripheral zone; TZ: transitional zone; SI: signal intensity; PCa: Prostate Cancer.*

Subgroup Analysis

Analyses made on subgroups of patients according to clinical, post-MRI biopsy and radiological parameters (Table 8), showed that k values were unreliable when the observations were unbalanced towards a specific category (i.e.: high prevalence of positive or negative MRIs), being disproportionately low compared to the raw percent

of agreement (PA). Conversely, AC1 provided stable results paralleling more faithfully PA values. Also in subgroup analyses, dedicated readers showed overall higher concordance than non-dedicated readers.

When accounting for clinical and radiological parameters, we observed higher agreement among all groups of readers in patients with PSA density (PSAD)≥0.15 ng/ml/cc, pre-MRI high and low risk of PCa, PZ lesions, homogeneous SI of the peripheral zone and easy interpretation of MRI, compared to patients with PSAD<0.15 ng/ml/cc, intermediate pre-MRI risk, TZ lesions, inhomogeneous SI of the peripheral zone and subjectively difficult interpretation of MRI, respectively (Table 8). Agreement was not significantly different in biopsy naïve or previous negative biopsy patients.

When accounting for post-MRI biopsy results, agreement was excellent in patients harboring csPCa (AC1=0.859, 95%CI: 0.782-0.936) and substantial in patients without PCa (AC1=0.744, 95%CI: 0.693-0.795). Conversely, concordance was significantly lower in patients with non-csPCa at biopsy (AC1=0.522, 95%CI: 0.344-0.701). Agreement in true positive findings of MRI was excellent (AC1=0.858, 95%CI: 0.780-0.936), while was low in false positive findings (AC1=0.528, 95%CI: 0.436-0.619). Among men who underwent biopsy, mean prevalence of csPCa in concordant positive cases across dedicated and non-dedicated readers was 52.3% (range: 48.1-56.5%), while in discordant cases was 10.7% (4.5-17.9%). Accordingly, mean false positive rate of MRI was 47,7% (43.5-51.9%) and 89.3% (82.1-95.5%) in concordant positive cases and discordant cases, respectively.

Concordance on the presence of more than one lesion at MRI was substantial (AC1=0.721; 95%CI: 0.603-0.838).

Indexes of Specific Agreement

Table 9 shows indexes of specific positive and negative agreement between dedicated and non-dedicated readers.

	Dedicated	Non-dedicated
Ppos (%)		
PI-RADS ≥3	75.7 (75.1-77.2)	63.6 (62.7-64.6)
PI-RADS 4-5	82.8 (81.2-84.3)	70.0 (68.6-71.4)
csPCa	93.4 (90.7-95.4)	86.0 (84.2-89.1)
Pneg (%)		
PI-RADS ≥3	85.1 (78.4-92.3)	82.0 (77.2-90.1)
PI-RADS 4-5	93.6 (90.5-96.5)	90.9 (87.3-94.5)
No PCa	87.8 (81.7-91.8)	86.0 (79.8-90.6)

Table 9. *Indexes of specific agreement of index lesion detection. Values in parenthesis are 95% confidence intervals. P_{pos}: proportion of positive agreement; P_{neg}: proportion of negative agreement*

When accounting for percentages of positive specific agreement, concordance was significantly higher for dedicated than for non-dedicated radiologists. Agreement on index lesion was substantial for a cut-off of PI-RADS score ≥ 3 and excellent for dedicated readers for a cut-off of PI-RADS score 4+5. Positive agreement was as high as 93.4% (95%CI: 90.7-95.4) between dedicated readers in patients with csPCa. Agreement of non-dedicated readers with a positivity threshold of PI-RADS score 4+5 approached that of dedicated readers with a positivity threshold PI-RADS score ≥ 3, even if it was still significantly lower.

Agreement on absence of lesions (negative MRI) was excellent both for dedicated and non-dedicated readers (respectively: 85.1%, 95%CI 78.4-92.3; 82.0%, 95%CI 77.2-90.1), and it was as high as 93.6 (95%CI: 90.5-96.5) for dedicated readers with a cut-off of PI-RADS score 4+5. In patients without PCa, negative specific agreement was excellent both for dedicated and non-dedicated readers (respectively: 87.8%, 95%CI 81.7-91.8; 86.0%, 95%CI 79.8-90.6).

Discussion

In our study we assessed the reproducibility of mpMRI reporting using PI-RADS v2.1 among multiple readers on a large cohort of patients who underwent mpMRI for a suspected PCa, reproducing a typical clinical workflow. We found overall good concordance among readers for index lesion detection, with excellent agreement in the subgroup of men harboring csPCa. As expected, concordance between experienced readers was generally higher than between less experienced readers.

Of note, agreement on absence of lesions was excellent across reader experience. To the best of our knowledge, this represents the first available study estimating the agreement on absence of lesions at MRI. This information is of paramount importance, given that the main strength of prostate MRI relies on its sensitivity and negative predictive value (Ahmed *et al*, 2017), and the most significant effect of its implementation has been to avoid biopsy in a substantial proportion of men (Mottet *et al*, 2019).

While agreement based on k values was comparable to previous multireader studies (Schimmöller *et al*, 2013; Muller *et al*, 2015; Rosenkrantz *et al*, 2016; Sonn *et al*, 2018; Pickersgill *et al*, 2018; Smith *et al*, 2019; Girometti *et al*, 2019), we confirmed how this statistical index may actually underestimate the true extent of the agreement, and how it could be unreliable in situations of unbalanced marginal frequencies compared to other coefficients (i.e.: AC1 coefficient) (Feinstein & Cicchetti, 1990; Cicchetti & Feinstein, 1990; Gwet, 2008). These evidences raise

questions about its suitability in prostate MRI image analysis, given that the probability of chance agreement in this setting is negligible (Shih *et al*, 2018). Correspondingly, our results are in line with other studies that evaluated indexes of specific agreement (Greer *et al*, 2017, 2019), and confirm that concordance in mpMRI reporting using PI-RADS scoring system may be actually higher than previously reported.

Besides readers' experience, other factors may affect the agreement. When positivity threshold of MRI was set to PI-RADS 4+5, variability has been significantly reduced; this is a clue of how PI-RADS score 3 lesions are a substantial source of interreader variability, especially in less experienced readers. Interestingly, to reach a similar level of positive agreement between experienced and non-experienced readers, the positivity threshold of MRI for less experienced readers should be heightened to PI-RADS scores 4+5 (Table 8). Furthermore, we observed more consistent scores between readers in patients at higher risk of PCa, for PZ lesions, in presence of homogeneous SI of the PZ and when MRI interpretation was subjectively judged easy (Table 8). In general terms, these findings may indicate that there is not one absolute value of interreader agreement: reproducibility is expected to be higher in those patients at higher risk of having an obvious lesion in the PZ, with less background PZ inhomogeneity, and thus in MRI examinations that are objectively easier to score. Accordingly, while agreement is higher for true positive findings, the majority of discordant cases (i.e.: harder to score cases) are more probably related to false positive findings. Interestingly, within the subgroup of patients who underwent biopsy, we observed that the false positivity rate of MRI was as high as 88,8% in presence of discordant findings between two different readers.

We also found good agreement on multifocality assessment, defined as the presence of more than one suspicious lesion (PI-RADS score $\geq$3) at MRI. However, taking into account the possibility of false positive agreement on non-index lesions, actual concordance on multifocal disease is expected to be lower, as previously reported (Greer *et al*, 2019). This consideration is relevant, as good concordance on multifocal disease in the presence of low lesion-specific agreement may furtherly support the need to perform random biopsies in adjunction to target biopsies in presence of suspicious lesions at MRI (Stabile *et al*, 2018; Rouvière *et al*, 2019).

Compared with other studies on interreader agreement, our study design has the advantage to reproduce a typical clinical scenario both in terms of image interpretation process and in patients' cohort, providing greater generalizability of the findings. Specifically, our readers had full access both to MRI examinations on a PACS workstation, as well as to pre-MRI clinical information; conversely, studies

based on scoring predetermined lesions on screen captures or blinding readers to relevant clinical information do not reflect the routine radiological workflow (Rosenkrantz *et al*, 2016; Muller *et al*, 2015; Schimmöller *et al*, 2013; Smith *et al*, 2019). Second, all available studies so far included only patients who underwent biopsy or radical prostatectomy (Schimmöller *et al*, 2013; Muller *et al*, 2015; Rosenkrantz *et al*, 2016; Pickersgill *et al*, 2018; Smith *et al*, 2019; Girometti *et al*, 2019; Greer *et al*, 2017, 2019), biasing the cohort toward high prevalence of positive MRI and clinically significant PCa. However, although useful for sub-analyses, histopathologic reference standard is not necessary for interobserver studies per se. Accordingly, we included all consecutive men who underwent MRI for clinically suspected PCa, regardless post-MRI biopsy results. This allowed us to have a cohort of patients that was more representative of the contemporary general population of men referred for prostate MRI for a clinical suspicion of PCa; at the same time, we were able to calculate agreement on negative MRI including patients that do not routinely undergo biopsy in clinical practice, according to latest guidelines (Mottet *et al*, 2019).

Our study is not devoid of limitations. First, despite our cohort was as much as possible representative of the general population, the retrospective nature of this study may have introduced selection biases. Therefore, prospective studies are needed to further validate our results.

Moreover, all readers came from a single tertiary care referral center, where less-experienced radiologists undergo specific training in prostate MRI led by experienced colleagues; this may induce readers to approach cases with similar interpretation schemes, reducing variability a priori. Thus, the generalizability of our findings may be limited to centers with a similar training. In the real world, the actual extent of variability between experienced radiologists from academic centers and non-experienced radiologists from non-academic centers may be significantly higher (Luzzago *et al*, 2019). Multi-center studies should be performed to address this limitation.

Finally, the lack of a direct comparison between PI-RADS v2 and v2.1 limited our possibility to assess the potential improvements in agreement that have been auspicated in the latest update. However, since PI-RADS v2.1 added minor changes to the previous version (some of which apply to relatively rare conditions, e.g.: central zone and anterior fibromuscular stromal tumors, atypical TZ nodules), to reliably detect significant differences between the two systems a larger number of cases is required. Future multicenter studies are needed to overcome this issue. Nonetheless, our study represents a valuable standpoint on agreement using PI-

RADS scoring system, giving insights on its clinical implications and on the investigative methodology that could be used for future studies.

In conclusion, we observed overall good reproducibility of prostate MRI interpretation between appropriately trained radiologists of different expertise using PI-RADS v2.1. Agreement is excellent between experienced readers in index lesion detection and across readers' experience in determining the absence of lesions at MRI.

Tackling interobserver variability in multiparametric MRI.

(The content of this chapter has been published by Brembilla et al. 2021, DOI: https://doi.org/10.1016/j.eururo.2020.10.023).

In the light of the results of our study (Brembilla *et al*, 2020a), and considering the studies available in the literature, some considerations can be made on the interobserver variability of mpMRI.

Whilst an MRI-influenced diagnostic pathway has shown advantages over traditional diagnostic methods (Kasivisvanathan *et al*, 2018), one of the key criticisms directed at prostate MRI is the level of agreement between different radiologists for identifying suspicious areas on a prostate MRI. Several studies have shown only low-to-moderate agreement between radiologists of various experience (Rosenkrantz *et al*, 2016; Pickersgill *et al*, 2018). Thus it is important for urologists to understand these studies in more detail to understand the implications to their practice and the implications when designing studies evaluating prostate MRI. Possible limitations in study design, patient selection and statistical approach can contribute to the observed results. We explain these, propose areas for future research and methods of overcoming these limitations in clinical practice.

To increase the generalizability of the results, interobserver studies should mirror clinical practice as closely as possible. Readers should interpret full examinations on imaging workstations rather than screen captures or annotated lesions, which has been seen in previous work (Rosenkrantz *et al*, 2016). The latter approach may be suitable for evaluation of specific imaging features within pre-specified lesions, but could underestimate interobserver agreement.

Similarly, the study population should be representative of that undergoing MRI in a clinical routine. However, biopsy or prostatectomy cohorts are frequently used, which biases the agreement due to the population having a higher prevalence of disease (Pickersgill *et al*, 2018). In fact, histopathologic examination is not required *per se* for interobserver studies, as they define the agreement on the presence or

absence of lesions at MRI regardless of the biopsy results. Moreover in using these cohorts, a lot of information regarding the agreement on the negative MRI is lost. Including patients with negative MRI is not always done but should therefore be encouraged (Brembilla *et al*, 2020a).

In addition, a lower observed agreement is possible due the statistical approach used for the analyses. K*appa* coefficient (κ) is generally accepted as the index of choice for interobserver studies. In the context of prostate cancer diagnosis, it gives the agreement between radiologists beyond chance for identifying suspicious lesions. However, it has some major drawbacks that need to be recognized and properly addressed. In particular, disproportionately low κ values can be observed despite high agreement when the observations are unbalanced towards one specific category (e.g. positive versus negative MRI) (Feinstein & Cicchetti, 1990). This situation is typically encountered in studies based on biopsy or prostatectomy cohorts. To address this issue, alternative coefficients could be used such as Gwet's agreement coefficient 1 (AC1) (Gwet, 2008), which has proven to be less prone to paradoxes than *kappa*. Moreover, non-chance-corrected indexes (e.g.: raw percentage of agreement and indexes of specific positive/negative agreement) can be informative in the setting of prostate MRI, since the probability that more readers detect the same lesion in the same location by chance is negligible (Brembilla *et al*, 2020a).

Accordingly, recent interobserver studies that were based on more balanced populations and used the correct statistical indexes reported very good agreement for detecting suspicious lesions with prostate MRI (Brembilla *et al*, 2020a; Greer *et al*, 2019).

Based on the current evidences, several ways of increasing the reproducibility of prostate MRI can be suggested. First, as MRI interpretation is subject to a learning curve (Garcia-Reyes *et al*, 2015), agreement is higher among more experienced radiologists and thus in clinical practice we would recommend that uro-radiologists with appropriate training should report prostate MRI (Rosenkrantz *et al*, 2016; Brembilla *et al*, 2020a). In institutions at the start of their MRI experience, we recommend working with another more experienced centre in order to reduce this learning curve. Second, standardized scoring systems (e.g. PI-RADS) can help reducing the differences in MRI interpretation and reporting (Brembilla *et al*, 2020a). These systems are being refined as we learn more about prostate MRI and thus interobserver agreement may continue to improve over time. Third, we would recommend that centres reporting prostate MRI should participate in multidisciplinary MRI and pathology review meetings which are commonly carried out in academic centres where interobserver agreement is higher (Rosenkrantz *et al*, 2016; Brembilla

et al, 2020a). This allows critical evaluation of each MRI interpretation relative to biopsy results so that the radiologists and urologists can identify areas for improvement.

However, little is known about the effect of MRI quality on MRI reproducibility and interobserver agreement. In fact, a standardized way of evaluating and reporting overall MRI quality in a study (Prostate Imaging Quality score, PI-QUAL) has been proposed only recently (Giganti *et al*, 2020). Research efforts should then be focused on reporting MRI quality as this could be a critical baseline measure in studies evaluating prostate MRI performance and reproducibility.

In summary, multiparametric MRI has established itself as a key tool in the prostate cancer pathway. Well-designed studies are required to reliably evaluate interobserver agreement of prostate MRI in clinical practice, which may help identify which factors leading to the highest agreement. When interpreting the result of interobserver studies of agreement, knowledge of the limitations of study design, patient selection and statistical measures of agreement are paramount for clinicians to understand in order to relate the findings to their own practice. When taking these limitations into account, we may find that MRI is even better than we think for prostate cancer diagnosis.

Diagnostic accuracy of abbreviated bi-parametric MRI (a-bpMRI) for prostate cancer detection and screening

(The content of this chapter has been published by Brembilla et al. 2022, DOI: https://doi.org/10.3390/ diagnostics12020231).

Multiparametric magnetic resonance imaging (mpMRI) plays a well-established role in the diagnostic workup of prostate cancer (PCa) (Kasivisvanathan *et al*, 2018). The standard mpMRI protocol consists of multi-planar T2-weighted imaging (T2WI), diffusion-weighted imaging (DWI) and dynamic contrast-enhanced (DCE) sequences (Turkbey *et al*, 2019). Evidence is accumulating that a shorter, non-contrast MRI protocol (i.e., bi-parametric MRI, bpMRI) may perform as well as mpMRI for the PCa detection (Kang *et al*, 2019; Jambor *et al*, 2019).

Recently, interest has grown over prostate MRI as a tool for PCa screening (Nam *et al*, 2016; Eklund *et al*, 2021; Eldred-Evans *et al*, 2021; Marsden *et al*, 2021a, 2021b). In this scenario, where shorter scan lengths would have important health economic implications, abbreviated bpMRI (a-bpMRI) protocols could further improve MRI accessibility and reduce costs (van der Leest *et al*, 2019a). However, evidence on their diagnostic performance is limited. Also, the optimal scoring approach for a-bpMRI protocols should be defined and tailored to the screening setting in terms of optimal cut-off scores and interpretation strategies (e.g.: single versus double reading) (Eldred-Evans *et al*, 2020).

The aim of this study is to investigate the diagnostic accuracy of a-bpMRI (<10 minutes scan) using template prostate mapping biopsies (TTPM) as the reference standard and to determine the best scoring approach to be used in a screening setting.

Materials & Methods

This study was conducted retrospectively on datasets acquired between 2012 and 2014 in the Prostate Imaging Compared to Transperineal Ultrasound-guided biopsy for significant prostate cancer Risk Evaluation (PICTURE) study (Simmons *et al*, 2017), a paired-cohort confirmatory study designed to assess the diagnostic accuracy of mpMRI in men undergoing TTPM and targeted biopsy. Ethical approval for the original study was granted by London City Road and Hampstead National research ethics committee REC reference 11/LO/1657.

Inclusion and exclusion criteria, along with the complete study protocol have been previously described (Simmons *et al*, 2014). Among the total participants to

PICTURE, 151 patients were included in the current study. Patient selection was carried out blinded to clinical data and MRI findings, with knowledge of biopsy results in order to obtain a predetermined prevalence of clinically significant cancer (i.e., 50% prevalence of Gleason 3+4 on TTPM biopsy), similar to that observed in another prospective, TTPM-based cohort of men with clinical suspicion of PCa (Ahmed *et al*, 2017).

Imaging protocols

All men underwent 3T mpMRI with a pelvic-phased array coil. The clinical mpMRI protocol consisted of the following sequences: axial and coronal T2WI, DWI including high b-value (b 2,000 s/mm^2) and apparent diffusion coefficient (ADC) map using multiple b-values (b 0, 150, 500, 1,000 s/mm^2), T1WI and dynamic contrast enhancement with gadolinium (Magnevist). Total duration of the clinical mpMRI protocol was 28 minutes and 14 seconds (Table 10). Sequence details are reported in supplementary Table 1.

Abbreviated bi-parametric MRI (a-bpMRI)

The a-bpMRI dataset was generated separately from the original mpMRI dataset: a-bpMRI included axial T2WI and b 2,000 s/mm^2 DWI sequences only (estimated duration: 9 minutes and 13 seconds) (Table 10).

Sequence	MpMRI	bpMRI	a-bpMRI
Localizer (T2WI – sagittal)	0:19	0:19	0:19
T2WI - axial	5:14	5:14	5:14
T2WI - coronal	5:55	5:55	-
DWI (b0, 150, 500, 1,000 s/mm^2)	5:17	5:17	-
DWI b 2,000 s/mm^2	3:40	3:40	3:40
T1WI	3:06	-	-
DCE	4:43	-	-
Total:	**28:14**	**20:25**	**9:13**

Table 10. *Protocol scan times (min:sec). Scan time of mpMRI and estimated scan times of bpMRI and a-bpMRI protocols. T2WI: T2-weighted imaging; DWI: diffusion weighted imaging; T1WI: T1-weighted imaging; DCE: dynamic contrast enhanced imaging; mpMRI: multiparametric MRI; a-bpMRI: abbreviated biparametric MRI.*

Image analysis

All scans were reviewed by three dedicated radiologists, all with more than 7 years of experience in prostate MRI reporting. Readers were blinded to clinical information, original mpMRI reports and biopsy results. The review process was two-fold: first, a session for a-bpMRI reporting; then a second session for bpMRI and mpMRI reporting, with an interval of 5 months between the two sessions. No feedback was provided to the readers until the end of the study.

1st session: abbreviated biparametric MRI

A-bpMRI scans were reported using a 5-point Likert scale for the likelihood of the presence of clinically significant cancer (1: highly unlikely; 2: unlikely; 3: equivocal; 4: likely; 5: highly likely) (Latifoltojar *et al*, 2019). Likert scales have been recommended for use in the UK by a consensus panel (Brizmohun Appayya *et al*, 2018) and recent UK National Institute for Health and Care Excellence (NICE) guidance (NICE Guidance - Prostate cancer: diagnosis and management: © NICE (2019) Prostate cancer: diagnosis and management, 2019), and were used for a-bpMRI scoring in the ReIMAGINE screening study (Marsden *et al*, 2021a). The following information were also collected: image quality of T2-WI and DWI sequences (i.e.: clear delineation of anatomical structures, no major artefact preventing sequence interpretation) and Likert scores for individual T2-WI/DWI sequences.

2nd session: biparametric and multiparametric MRI

During the second session readers scored the bpMRI images first, then the full protocol (including DCE) was revealed for mpMRI scoring. BpMRI and mpMRI scans were reported using both Likert and PI-RADS v 2.1 score.

Histopathologic assessment

All men underwent TTPM biopsies regardless the MRI results within the PICTURE study (Simmons *et al*, 2017). Biopsies were performed using a brachytherapy grid fixed on a stepper with cores taken every 5 mm. All biopsies were reported by one of two uro-pathologists with more than 20 years of experience, blinded to the mpMRI reports.

Clinically significant cancer was defined by criteria developed and validated for TTPM biopsies (Ahmed *et al*, 2011). The primary definition (definition 1) of csPCa was considered the presence of dominant Gleason pattern 4 or greater (i.e., Gleason ≥ 4+3) or a cancer core length (CCL) involvement of ≥ 6mm of any Gleason score.

Alternative definitions were: i) Gleason ≥3+4 and/or CCL ≥4 mm (definition 2) and ii) presence of any Gleason ≥7.

Statistical analyses

Contingency tables were used to calculate the diagnostic accuracy of the tests on a patient level for a cut off of MRI score ≥ 3: sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and MRI positivity rate (i.e., proportion of men with positive MRI finding according to cut-off score) were reported with 95% confidence intervals. True positive, false positive, true negative and false negative results from each reader were combined by simple pooling to calculate diagnostic accuracy measures. Differences in sensitivity and specificity between a-bpMRI, bpMRI and mpMRI were assessed using Mc Nemar's test, while differences in predictive values were assessed using a general estimating equation logistic regression model (Leisenring *et al*, 2000; Wang *et al*, 2006).

The diagnostic accuracy of a-bpMRI was then tested using alternative cut-offs score: 1) overall MRI score ≥ 4; 2) presence of any lesion with both T2WI and DWI scores ≥4. A-bpMRI scores from each reader were evaluated both individually and in combination (referred to as "combined scores") to simulate a double reading approach in which two radiologists report the independently with a third reviewer involved when there is disagreement.

Finally, sensitivity and specificity values of a-bpMRI were used to predict the variation of PPV, NPV and MRI positivity rates (i.e. the proportion of men with a positive MRI over the total of men undergoing MRI) for lower prevalence of csPCa (2%, 5% and 10%) (Altman & Bland, 1994).

Interobserver agreement was assessed using Gwet's agreement coefficient 1 (AC1) and percentage of agreement (PA) (Gwet, 2008; Cicchetti & Feinstein, 1990). Levels of agreement were interpreted using the classification of Landis and Koch (Landis & Koch, 1977).

Statistical tests were performed using RStudio graphical interface for R software v.4.0.2 and AgreeStat v.2015.6.

Results

Baseline characteristics

A summary of baseline demographic data of men included in this study is presented in Table 11. Median age was 62 years (range: 41-83), median PSA was 6.8 ng/ml (range: 0.9-28.5). Prevalence of clinically significant cancer at TTPM biopsy

was 60/151 (40%) according to definition 1, 95/151 (63%) according to definition 2 and 76/151 (50%) for any Gleason ≥7.

No. of Patients	151
Median age, y (range)	62 (41-83)
Median PSA, ng/ml (range)	6.8 (0.9-28.5)
Highest Gleason grade at Histopathology (%)	
Benign	22 (15)
3+3	53 (35)
3+4	63 (42)
≥ 4+3	13 (8)
Definition 1* csPCa	60 (40)
Definition 2** csPCa	95 (63)
Any Gleason ≥ 3+4	76 (50)
Median no. of positive cores (IQR)	6 (2-11)
Median MCCL (IQR)	4 (1.5-7)

***Table 11**. Summary of demographic data. * Gleason ≥4+3 and/or cancer core length (CCL) involvement ≥ 6mm. ** Gleason ≥3+4 and/or CCL ≥4 mm. csPCa: clinically significant prostate cancer; IQR: interquartile range; MCCL: maximum cancer core length.*

Diagnostic accuracy of MRI

On a-bpMRI, axial T2WI sequences were judged of good diagnostic quality by all readers for all patients; DWI sequences were judged of low diagnostic quality in 10, 5 and 11 patients by the three readers respectively due to the presence of image artefacts.

Sensitivity, specificity, PPV and NPV of a-bpMRI with a cut-off of MRI score ≥3 for the detection of definition 1 csPCa were 92% (95%-CI: 87-96), 48% (95%-CI 42-54), 54% (95%-CI: 48-60) and 90% (95%-CI: 84-95), respectively; there was no significant difference in sensitivity and NPV between a-bpMRI, bpMRI and mpMRI (Table 12). The diagnostic accuracy of MRI for alternative definitions of csPCa is reported in Supplementary Table 2.

The proportions of MRI scores (1-2, 3 and 4-5) were similar for a-bpMRI, bpMRI and mpMRI PI-RADS scores (32%, 14% and 54% VS 37%, 13% and 50% VS 36%, 14% and 50%, respectively). A higher proportion of MRI score 3 was observed for bpMRI-mpMRI when using Likert scores (up to 7% increase) (Supplementary Figure 1).

	a-bpMRI	**bpMRI**		**mpMRI**	
	Likert	**Likert**	**PI-RADS**	**Likert**	**PI-RADS**
Sensitivity	92 (87-96)	92 (87-96)	89 (83-93)	92 (87-96)	89 (83-93)
Specificity	48 (42-54)*	35 (29-41)	52 (46-58)	39 (33-45)	53 (47-59)
PPV	54 (48-60)*	48 (43-54)	56 (49-61)	50 (44-56)	56 (50-61)
NPV	90 (84-95)	87 (79-93)	88 (82-92)	88 (81-94)	88 (82-92)

***Table 12**. Diagnostic accuracy of MRI for definition 1 csPCa. Pooled values are reported as % (95%-CI). *p<0.05 a-bpMRI vs bpMRI/mpMRI Likert. bpMRI: biparametric MRI; mpMRI: multiparametric MRI; a-bpMRI: abbreviated biparametric MRI; PPV: positive predictive value; NPV: negative predictive value.*

A-bpMRI: alternative cut-offs, combined scores and csPCa prevalence

Higher MRI cut-offs (MRI score ≥4 or higher) were associated with lower sensitivity (range: 70-83%) and higher specificity (range: 64-76%) for csPCa detection (Table 13; Supplementary Table 3); combined scores yielded slightly better performance, even if the difference was not significant.

	MRI score ≥4	T2WI and DWI score ≥4
Pooled		
Sensitivity	83 (76-88)	70 (63-77)
Specificity	64 (58-69)	76 (71-81)
PPV	60 (54-66)	66 (59-73)
NPV	85 (79-89)	79 (74-84)
Combined*		
Sensitivity	85 (76-94)	72 (60-83)
Specificity	65 (55-75)	79 (71-87)

PPV	61 (51-72)	69 (58-81)
NPV	87 (79-95)	81 (73-89)

Table 13. *Pooled and combined diagnostic accuracy of a-bpMRI according to different MRI cut-offs (for definition 1 csPCa). Pooled values (aggregate from the three readers) are reported as % (95%-CI). *Simulated double reading approach: two radiologists report the a-bpMRI independently with a third reviewer involved when there is disagreement between the reporters. MCCL: maximum cancer core length; PPV: positive predictive value; NPV: negative predictive value.*

Assuming sensitivity and specificity of MRI remain constant, predictive values were simulated for variable levels of csPCa prevalence (10, 5 and 2%). At lower levels of prevalence, PPV decreased while NPV increased: for an hypothetic csPCa prevalence of 5% as described in screening populations (Thompson *et al*, 2004), NPV was very high (range: 98-100%) regardless the MRI cut-off used (Table 14; Supplementary Table 4). Higher MRI cut-offs were associated to lower positivity rates of MRI (Table 5).

	Prevalence of csPCa (definition 1)		
	10%	**5%**	**2%**
MRI score ≥3			
Pos.rate	59 (51-67)	57 (49-65)	56 (47-64)
PPV	16 (9-25)	8 (3-16)	4 (1-10)
NPV	98 (91-100)	100 (94-100)	100 (95-100)
MRI score ≥4			
Pos. rate	40 (32-49)	37 (30-46)	36 (29-45)
PPV	21 (12-34)	11 (4-22)	5 (1-15)
NPV	98 (92-100)	99 (94-100)	100 (96-100)
T2WI and DWI score ≥4			
Pos. rate	26 (19-34)	23 (17-31)	22 (16-29)
PPV	28 (15-45)	14 (5-30)	6 (1-20)
NPV	96 (91-99)	98 (94-100)	99 (95-100)

Table 14*. PPV, NPV and positivity rates of abbreviated bpMRI (a-bpMRI) according to clinically significant PCa prevalence. Values are reported as % (95%-CI). Pos. rate: rate of positive test according to MRI cut-off. csPCa: clinically significant prostate cancer; pos. rate: positivity rate of MRI; PPV: positive predictive value; NPV: negative predictive value.*

Missed csPCa

Most of the lesions missed by mpMRI and a-bpMRI were of low-to-intermediate grade (Gleason 3+3 or 3+4), while one Gleason 4+3 tumour was missed by all MRI

protocols regardless the MRI cut-off (Supplementary Table 5). The number of missed high grade tumours did not increase with a-bpMRI compared to mpMRI. Conversely, it increased when a cut-off of MRI score ≥4 on both T2W and DWI was used for a-bpMRI.

Interreader agreement

Interreader agreement of a-bpMRI was moderate and lower than mpMRI PI-RADS scores (PA: 76% VS 81%, respectively; AC1 0.58 VS 0.65), while it was comparable to that of bpMRI (Supplementary Table 6). Agreement on a-bpMRI scores increased with higher positivity thresholds and was as high as 82% (CI-95%: 77-87; AC1 0.64, CI-95% 0.53-0.72) for a cut-off of MRI score ≥4 on both T2WI and DWI (Supplementary Table 7).

Discussion

The present study provided evidence that an abbreviated biparametric MRI protocol consisting in axial T2WI and high b-value DWI sequences only (<10 min scan time) could match the performance of full biparametric and multiparametric MRI protocols for the detection of csPCa. Our findings support emerging evidence on the limited utility of DCE for PCa detection (Bosaily *et al*, 2020; Bass *et al*, 2020) and are in line with similar studies that investigated abbreviated bpMRI protocols (van der Leest *et al*, 2019b; Kuhl *et al*, 2017; Obmann *et al*, 2018; Barth *et al*, 2017; Weiss *et al*, 2018; Cereser *et al*, 2020).

Fast MRI protocols that can be performed in less than 10 minutes could have a favourable impact on costs and accessibility of MRI (van der Leest *et al*, 2019a). As a consequence, bpMRI has recently drawn attention as a potential tool for imaging-based PCa screening, comparable to mammography for breast cancer or low-dose CT scan for lung cancer (Nam *et al*, 2016; Eklund *et al*, 2021; Eldred-Evans *et al*, 2021; Marsden *et al*, 2021a, 2021b). However, the screening setting is different from that of secondary care where prostate MRI was developed, and little is known about the impact on MRI performance: notably, the prevalence of csPCa in the general population is significantly lower than in men with a clinical suspicion of PCa (Thompson *et al*, 2004). In this setting, attention should be focused on ruling out significant PCa while minimizing the number of positive MRI to avoid unnecessary interventions (Schoots *et al*, 2021). One of the most straightforward way to achieve this goal is to use higher thresholds for MRI positivity. Accordingly, we simulated the performance of a-bpMRI at lower prevalence levels and we observed that higher cut-

offs (e.g., MRI score ≥4 or higher) yield a more acceptable balance between sensitivity and specificity, significantly reducing the proportion of positive scans and false positive findings. At the same time, the NPV is expected to remain high (>97%) due to the low prevalence of the disease.

These numbers compare favourably with other established screening modalities: for lung cancer screening using low-dose CT, positivity rates of 27% and overall PPV of approximately 4% have been reported (The National Lung Screening Trial Research Team, 2013). However, our findings should not be considered an exact prediction of the accuracy of MRI in a screening setting. Rather, they are intended to provide an insight of the effect that different MRI cut off, scoring approach and disease prevalence would have on the ability to predict the presence/absence of csPCa, in order to guide future applications of MRI for PCa screening. Real numbers are likely to be affected by true prevalence (which will vary with age and patient selection), prior tests, patient setting, scoring systems, biopsy techniques and definitions of csPCa. In the recently published IP1-PROSTAGRAM study where MRI was used as a primary screening test (Eldred-Evans *et al*, 2021), a total of 17 clinically significant cancers were found in 406 men screened (observed prevalence: 4.2%). Forty-three out of 406 men had a positive MRI using a cut-off of PI-RADS score 4-5 (scan positive rate: 10.6%); among these, csPCa was found in 11/43 men (PPV: 25.6%). Interestingly, these data overlap with those predicted in our study for a 5% prevalence of Definition 2 csPCa (supplementary Table 4).

One of the major drawbacks that has been associated with the use of abbreviated MRI protocols is the potential increase of equivocal findings (i.e. PI-RADS/Likert score = 3) (van der Leest *et al*, 2019a; Zawaideh *et al*, 2020). In our study, a-bpMRI did not lead to an increased number of equivocal findings compared to standard bpMRI and mpMRI; moreover, the impact of equivocal scores would be negligible if a cut off of MRI score ≥ 4 is used (Eldred-Evans *et al*, 2021). Conversely, we observed lower interobserver agreement of a-bpMRI compared to previous reports (van der Leest *et al*, 2019a; Brembilla *et al*, 2020a). One possible explanation is that T2WI, DWI and DCE have an incremental value on lesion visibility, and the absence of multiplanar T2WI and ADC maps from our a-bpMRI protocol may have decreased the confidence of readers in determining MRI scores (Huebner *et al*, 2020). Furthermore, the use of Likert instead of PI-RADS for a-bpMRI could have played a role (Supplementary Table 6) (Brembilla *et al*, 2020b).

In our study, we also simulated the performance of a double reading approach (i.e. two radiologists report the MRI independently with a third reviewer involved in

case of disagreement), that replicates the methodology used for screening mammography (Thurfjell *et al*, 1994). While we did not observe any significant benefit (nor disadvantage) in terms of diagnostic performance of combined VS individual reads, this approach could be still considered as a potential solution to reduce the variability in MRI interpretation of abbreviated protocols. Moreover, second reads could represent a future application of artificial intelligence tools (Penzkofer *et al*, 2021).

One of the main strengths of our study is that all patients underwent mapping biopsies regardless the MRI results, increasing the reliability of sensitivity and specificity estimates compared to conventional targeted biopsy cohorts. Second, the multireader design averages different readers' performance that can be encountered in the clinical routine, increasing the generalizability of the results. Third, a-bpMRI and bp/mpMRI were reported in two different sessions separated by a wash-out period. This may enhance the detection of small differences in the diagnostic accuracy between different study protocols when compared to standard sequential reading, where the results of one test will be invariably influenced by the other.

This study has also several limitations. First, patients were selected from a broader cohort of men included in a prospective trial (Simmons *et al*, 2017), potentially introducing a selection bias. However, the demographics and csPCa prevalence were similar to those of other studies based on unselected populations (Ahmed *et al*, 2017), supporting the generalizability of our cohort.

Second, although we did not find significant differences in the performance of a-bpMRI, bpMRI and mpMRI, the study was not specifically powered to detect such small differences and our findings need to be confirmed by larger studies. Third, Likert scores were used for a-bpMRI reporting, limiting the reproducibility of our results and the comparison with PI-RADS scores. However, although PI-RADS can be applied to bpMRI, this scoring system has not been conceived for short protocols (i.e.: monoplanar, no ADC maps) (Turkbey *et al*, 2019). Fourth, all the readers involved in the study are experienced readers from high volume centres, a factor that limits the applicability of our results to low volume settings and less experienced readers.

In conclusion, abbreviated biparametric MRI (<10 min scan) can be as accurate as mpMRI for the detection of csPCa in the diagnostic setting when interpreted by expert readers. As far as scan time, availability, cost and acceptability are concerned, one might argue that an a-bpMRI could be the ideal tool for imaging-based screening programs for PCa. If used for this purpose, higher cut offs (MRI score ≥4) could yield a more favourable balance between PCa detection and false positive rates and could

be preferred over lower cut-offs (MRI score ≥3). A double reading approach might be considered to address the variability of MRI interpretation.

Shortened Luminal Index MR imaging (LI-MRI) for the detection and characterization of significant PCa

The prostate gland is predominantly composed by three components: luminal space, epithelial and stromal cells (Figure 2). In the luminal spaces is stored the fluid produced by the glandular epithelial cells, while the stromal cells form a scaffold that gives the prostate a solid structure, also helping directing the prostatic fluid out of the prostate during ejaculation.

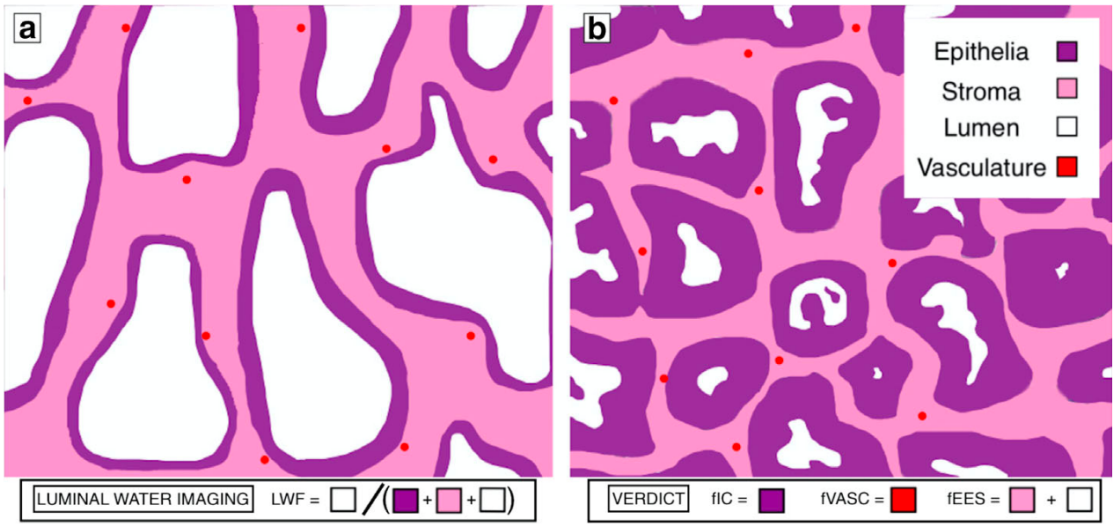


Figure 2. *Histologic composition of the prostate gland (Devine et al, 2019).*

It is known from histology studies that cancer, specifically Gleason 3+4 or higher, alters the balance of glandular to cellular spaces, reducing the fraction of glandular lumen. This fraction decreases further as the Gleason grade of tumour increases. Furthermore, as already described in the Introduction chapter, the prostate is made of different tissue zones: peripheral, transition and central zones. Each is composed of different proportions of gland, epithelium and stroma. For example, the central zone consists of large irregular glands with cuboidal epithelial cells and compact stromal tissue. The peripheral zone contains small regular glands with columnar epithelial cells and is surrounded by loose stroma. The normal transition zone resembles the peripheral zone tissue, but it undergoes dramatic changes with advancing age (benign prostatic hyperplasia, BPH).

Thus, MRI techniques that can distinguish between those component may help characterizing PCa, adding biologic specificity to the standard mpMRI that is known to be associated with poor specificity (Drost et al, 2019). Specifically, DWI that is the most important sequence in mpMRI and highlight area of restricted water diffusion

that are usually referred to as areas of increased cellularity. However, restricted diffusion may be related to alternative biologic process and also benign entities (such as inflammation) may cause a focally increased prostatic cellularity.

Two MRI techniques have been described that can characterize the prostate gland microstructure. The vascular, extracellular, and restricted diffusion for cytometry in tumors (VERDICT) technique combines a diffusion-weighted MRI acquisition with a mathematical model and assigns the diffusion-weighted MRI signal to three principal components: (a) intracellular water, (b) water in the extracellular extravascular space, and (c) water in the microvasculature (Johnston et al, 2019). The luminal water imaging (LWI, also called Luminal Index MRI – LI-MRI) allows the modelling of two compartments within the prostate gland: (a) the luminal space, composed of a distribution of long T2 values; (b) the stromal + epithelial spaces, composed of a distribution of short T2 values (Devine et al, 2019).

Evidence has accumulated in the literature on the performance of multiecho T2 (ME-T2) sequences. First, ME-T2 sequences have been described to reflect prostate tissue composition better than conventional T2W sequences, since the mono-exponential fit is able to adequately depict tissue microstructure in only approximately 10% of men (Storås et al, 2008). Subsequently the LWI technique has been proposed as a method to quantify the T2 distribution in the prostate tissues (Sabouri et al, 2017); specifically, LWI allows the estimation of the of the luminal water fraction (LWF), that is the fraction of the luminal water component in each voxel (vs epithelial/stromal components). Of note, the MRI-derived LWF has a good correlation with the histologically measured luminal fractional volume, harbouring the potential to determine the presence of PCa and predict Gleason score (Sabouri et al, 2017). However, the original LWI technique employed non-standard MRI sequences (T2 weighted multi-echo acquisition with 64 echoes) which required complex set-up and provided one image slice taken through the whole prostate.

At UCL, Devine et al. (Devine et al, 2019) further investigated the applicability and performance of a simplified LI-MRI protocol based on 32-echoes ME-T2 sequences that sampled the whole prostate. The results were encouraging, showing that a simplified LI-MRI protocol was feasible and that LWI alone predicted PI-RADS scores and detect csPCa with similar performance compared to ADC alone. Also, LWI also provides a more specific multicompartment information that has an explicit biophysical correlation.

Still, the 32-echo sequences have several limitations, such as a reduced spatial resolution, a partial sampling of the prostate volume and long scan times. To address these issues, in the present project we investigate a further refinement of the LI-MRI

technique (i.e. further reduction of the ME-T2 echoes with whole gland coverage and increased spatial resolution) that could pave the way to its implementation in the clinical practice.

A two stage study was conducted: first, the diagnostic performance of 32-echoes versus 8-choes derived LWF was compared; second, repeatability and proof-of-concept of a shortened 8-echo full prostate coverage luminal water imaging (LWI) sequence was performed.

Materials & methods

Patient Cohort

Participants were recruited into INNOVATE study (CombIning advaNces in imagiNg With biOmarkers for improVed Diagnosis of Aggressive prosTate cancEr) and gave written informed consent for additional MRI sequences. INNOVATE is registered at ClinicalTrials.gov (NCT02689271) and Research Ethic Committee reference number is 15/LO/0692. Men undergoing investigation for suspected prostate cancer were identified from a single tertiary referral centre between July 2016 and December 2019. Inclusion criteria were: age ≥ 18y and referral for prostate mpMRI based on clinical suspicion of prostate cancer. Exclusion criteria were: inability to undergo full mpMRI, bilateral hip replacements; inability to give informed consent; previous treatment (prostatectomy, radiotherapy, brachytherapy or focal therapy) for prostate cancer; on-going hormonal treatment for prostate cancer, and previous prostate biopsy within 6 months of scheduled mpMRI.

The whole cohort was split into 212 men for stage 1 (comparison of 32-echo and 8-echo model fitting approaches), and 39 men for stage 2 (repeatability and proof-of-concept of a full prostate coverage 8-echo LWI pulse sequence). Study flowchart is shown in Figure 3; participant demographic data are shown in Table 15.

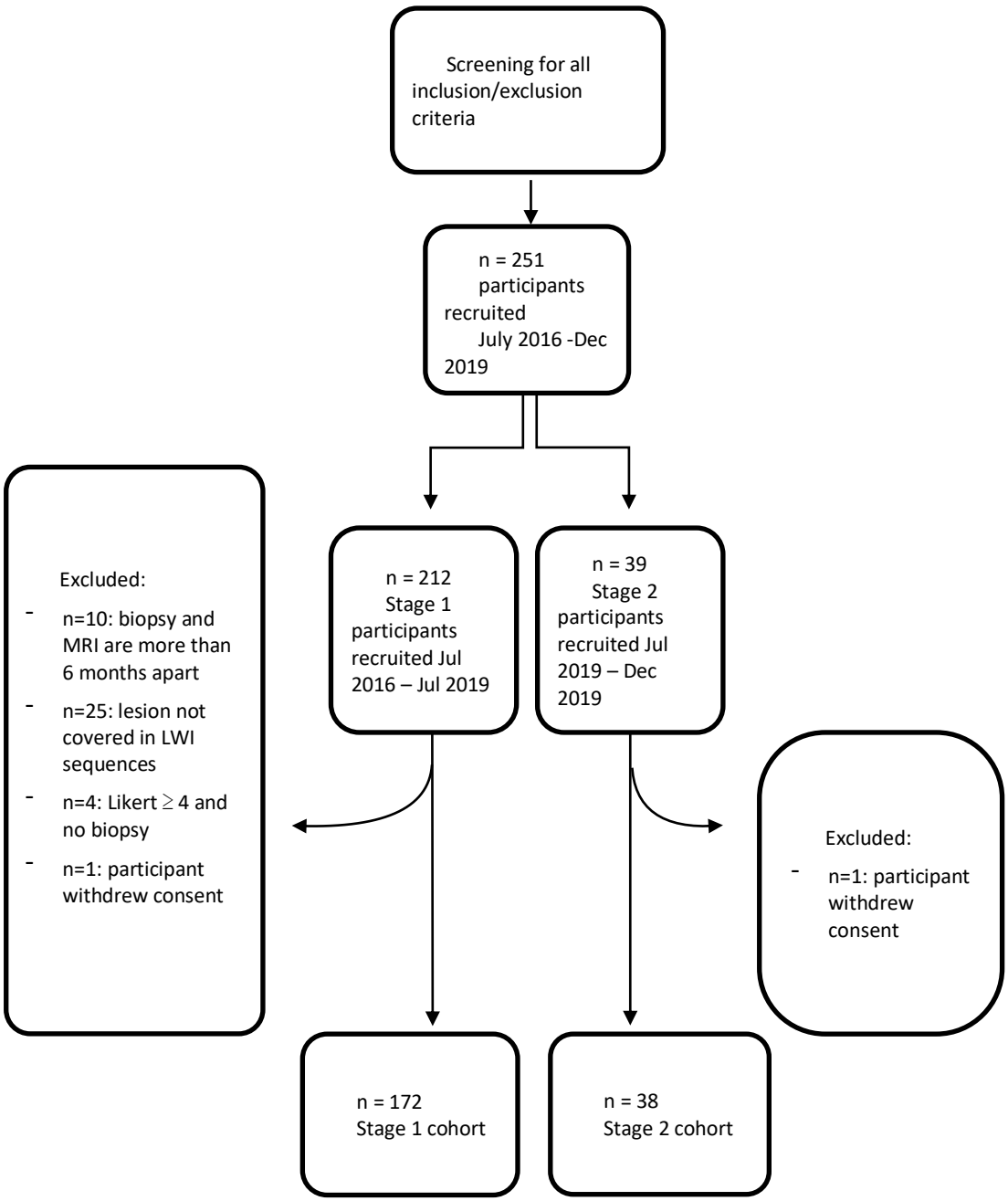


Figure 3. *Study flowchart.*

Parameter	Whole Cohort	Stage 1	Stage 2
No. Of Participants	210	172	38
Median Age (y)	64 (40-80)	(49-80)	64.5 (44-80)
Median PSAd (ng/ml)	0.12 (0.02-2.39)	0.13 (0.03-2.39)	0.1 (0.02-2.07)

Maximum Likert score			
2	34	28	6
3	101	81	20
4	41	33	8
5	34	30	4
Highest Gleason grade of biopsied index lesion			
Negative	57	48	10
3+3	13	13	0
3+4	33	25	8
> 3+4	24	23	1
TOTAL Biopsied	127	109	19

Table 15. *Participant demographic data.*

Image acquisition

Imaging was performed on a 3.0T scanner (Philips Achieva; Philips Medical Systems, Best, The Netherlands) using a 32-channel cardiac coil. All participants underwent a mpMRI as part of standard of care. In addition, either at the time of the mpMRI or during a second hospital visit (within a median of 2 days) following the mpMRI, the stage 1 cohort were scanned with a 32-echo luminal water sequence (limited to 6 axial slices and a 2x2x4 mm resolution; Table 16).

Stage 2 patients were invited to undergo 8-echo LWI sequence of the prostate, that was repeated twice within the same session to assess repeatability of derived LWF measurements.

	Stage 1	**Stage 2**
Parameter	32-echo	8-echo
Number of echoes	32	8
Echo Spacing (ms)	31.25	31.25
TR (ms)	8956	7675
Acquisition voxel size (mm^3)	2x2x4	2x2x4
FOV (mm^3)	180x180x26	180x180x68
Number of slices	6	17

Slice gap	0.4mm	0.4mm
Scan Duration (mm:ss)	05:49	02:56

Table 16. *MRI parameters. Participant in Stage 1 had a single 32-echo sequence, and Stage 2 participants had 2 8-echo sequence each repeating twice back to back.*

The mpMRI study was reported as standard of care by board-certified radiologists with ≥10 years of prostate MRI experience. Reporting was performed as per the UK NICE guidelines using Likert scoring to highlight lesions for biopsy consideration. Radiologists were blinded to the LWI.

Image analysis

For the purposes of LWF analysis, the index lesion was identified on mpMRI as the highest Likert score lesion. In cases of multiple lesions with the same Likert score, the largest lesion was chosen. An experienced radiologist with 7 years of mpMRI reporting experience, aware of the index lesion sites and of targeted biopsies but not aware of the biopsy outcome, contoured lesions on the T2 weighted images of the mpMRI dataset. The radiologist then transferred the mpMRI T2 weighted contoured lesion to the corresponding LWI slice for calculation of lesion LWF. For participants with a maximum Likert score of 1-2, as no lesion was present, the radiologist placed a single circular region of interest (ROI) within the mid-gland peripheral zone to derive LWF. In addition, all index lesions were rescored using PIRADS v2.1 by the study radiologist.

For repeatability analysis (stage 2), a study radiologist contoured the left and right transition zone and the left and right peripheral zone on every slice containing prostate tissue on the 3rd echo (TE 93.75ms) of the LWI images of the first acquisition. ROIs were then copied onto the second acquisition and finally onto the corresponding LWF maps. This creates a total of 914 matched ROIs.

Luminal Water Fraction calculation

For the stage 1 cohort, two sets of LWF maps were generated with the method proposed by Devine et al. using: (i) all 32-echoes acquired; and (ii) only the first 8-echoes of each 32-echo acquisition. For stage 2 cohort, 1 set of LWF maps were calculated using the 8 echoes of the native sequence.

The Devine method consists in fitting the multi-echo T2 images to a simplified model which uses two Gaussian distributions to simulate the T2 decay curve. This fitting model minimises the mean square error between actual signal and simulated signal over six parameters: M_0 (absolute signal magnitude), α (the magnitude ratio

between two peaks), μ1(mean of short T2 peak), μ2(mean of long T2 peak), □1 (variance of short T2 peak) and σ_1 (variance of long T2 peak). μ1 represent the compartment composed of stroma and epithelia which has shorter T2 value and μ2 represent the luminal compartment with longer T2 values. The values of μ1 and μ2 are constrained to be 0-200ms and 200-2000ms respectively. LWF is calculated for each pixel as area under long T2 distribution divided by the sum of area under short and long T2 distribution. The area is computed by integrating the respective Gaussians using their magnitude, mean, and variance.

The median LWF of ROIs was calculated for the 32-echo and 8-echo derived LWF maps respectively. All data was processed using MATLAB [R2020b (9.9.0.1524771)].

Standard of reference

MpMRI-targeted biopsies were used as the standard of reference. Definition of clinically significant cancer (csPCa) was Gleason 3+4 disease. Patients harbouring csPCa were defined as "significant" group; those with negative biopsy or Gleason 3+3 cancer were classified as non-significant group ("NS"). For participants with a Likert 2-3/5 score where biopsy was not performed, LWF ROI values were assigned to the NS group, given the low likelihood of harbouring csPCa.

Statistical analysis

A Wilcoxon matched pairs test was performed to determine if there was a difference between LWF value generated from 32-echo and 8-echo respectively across all participants (n=172).

A Mann-Whitney comparison was performed to determine differences in LWF between 'significant' and 'NS' groups (n=172) for both 32-echo and 8-echo LWF calculations. Similarly, classification performance was evaluated by receiver operating characteristic (ROC) area under curve (AUC) analysis for both 32-echo and 8-echo LWF derivations.

The Kruskal-Wallis test was used to compare LWF values across individual Likert scores (n=172) and across mpMRI biopsied lesion Gleason grades (n=109) for the 32-echo and 8-echo calculated LWF values respectively.

A sub analysis was performed (for 32 and 8-echo LWF respectively) including only lesions with known histological outcome (n=109). LWF was compared between 'significant' and 'NS' groups and across Likert 3, 4 and 5 scores; and ROC-AUC analysis was repeated. Finally, a second sub-analysis was performed for biopsied Likert 3-4 scored lesions only (n=79 for Likert, n=73 for PIRADS).

A threshold value of LWF was derived from the 8-echo ROC-AUC for all participants (n=172) at 90% sensitivity for subsequent application to the Stage 2 proof-of-concept cohort. The above analysis were repeated with PIRADS scores.

Bland Altman analysis was conducted to determine bias and 95% limits of agreement between repeated acquisitions. Sensitivity and specificity of the LWF threshold for significant cancer was calculated for biopsied lesions.

Data were analysed using PRISM [Version 8.3.0].

Results

Stage 1: Comparison of performance of 32-echo and 8-echo diagnostic performance.

Across all participants (n=172), LWF values were significantly lower when calculated using 8-echoes compared to 32-echoes (Figure 4; median LWF 8-echo: 0.095 vs. 32-echo:0.156; p<0.001). To differentiate results from these two datasets and to acknowledge that the measurement is not identical, from here on in the 8-echo derived LWF will be termed the Luminal Index (LI).

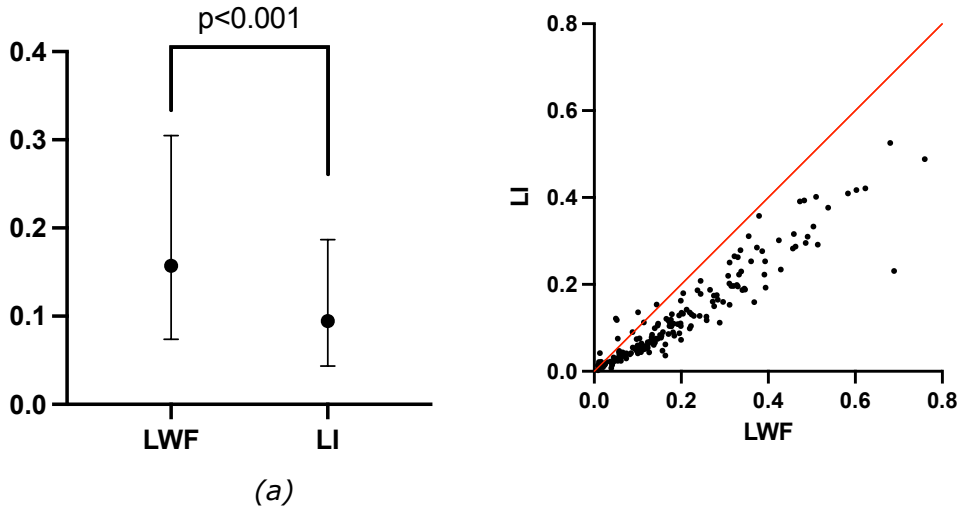


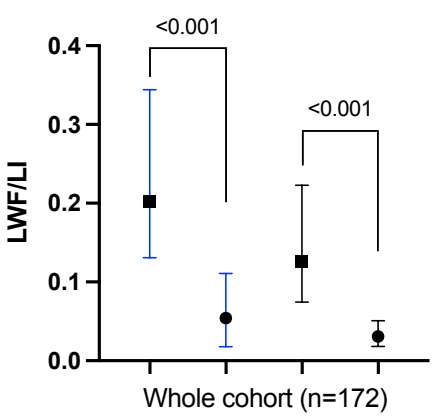
Figure 4. *LWF and LI values compared. Wilcoxon matched pair test (n=172) between LWF (derived from 32-echo sequence) and LI (derived from 8-echo sequence). (a) shows median with interquartile range (IQR); (b) shows X-Y plot of the matched LWF and LI values.*

Across all participants (n=172), LWF was significantly lower (Mann-Whitney) in the 'significant' group compared with the 'NS' group (median LWF 0.054 vs. 0.202, p<0.001) (Figure 5a). This significant difference was retained when LI was compared between the two groups (median LI 0.031 vs. 0.127, p<0.001).

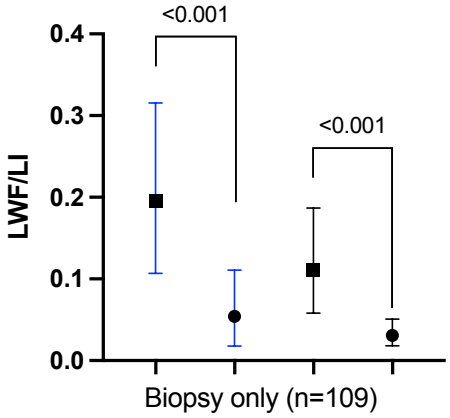
The ROC-AUC for LWF and LI classification of significant cancer was 0.87 and 0.88 respectively. A LI value of 0.095 provided 90% sensitivity and 65% specificity for significant cancer.

Of the 109 index lesions biopsied, 48 were negative for cancer (44%), 13 were Gleason 3+3 (12%), 25 were Gleason 3+4 (23%) and 23 (21%) were Gleason >3+4 disease; LWF and LI were both significantly lower in the 'significant' group (n=48) compared with their respective values in the 'NS' group (n=61), with p<0.001 (Figure 5b). The ROC-AUC for LWF and LI classification of significant cancer was 0.84 and 0.85 respectively. The same LI value threshold as found for the full cohort analysis (LI=0.095) provided a 90% sensitivity but a more limited specficiity 56% for significant cancer.

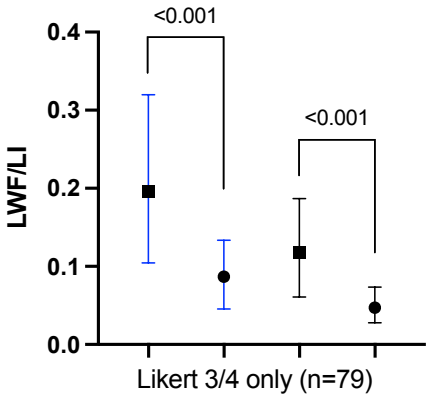
Across Likert 3 or 4 scored biopsied participants only (Figure 5c, n=79), LWF and LI were significantly lower in the 'significant cancer' versus 'NS' group (median LWF 0.08685 vs 0.1972; p<0.001; median LI 0.04725 vs 0.1181; p<0.001). The ROC-AUC for LWF and LI classification of significant cancer was 0.79 and 0.80 respectively.



(a)



(b)



(c)

Figure 5. *LWF and LI-values in NS and significant groups. Mann-Whitney test for LWF and LI across: All participants (n=172) in figure (a), biopsied participants only (n=109) in figure (b), and Biopsied Likert 3/4 only cohort (n=79) in (c). In all subfigures, blue lines denote LWF and black are for LI, solid squares are the median values for NS group and solid dots are median value for significant group in perspective subgroups.*

Figure 6 illustrates difference in LWF and LI for the whole cohort (n=172) across Likert scores. Median LWF and LI reduced with increasing Likert score. However differences in LWF and LI between Likert 3 and 4 groups were not statistically significant (p=0.22 and p=0.07 respectively).

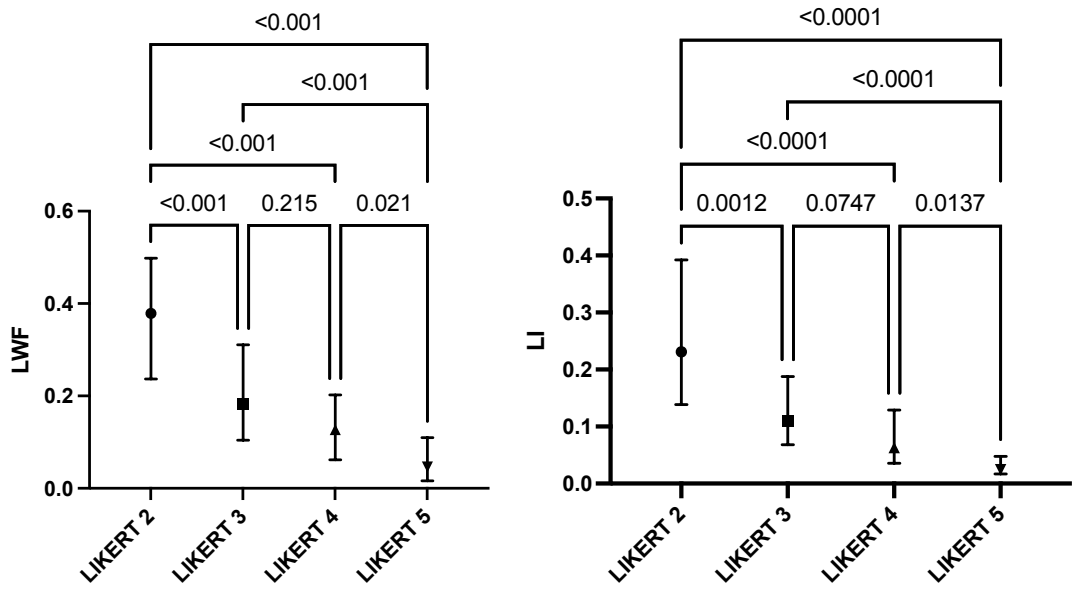


Figure 6. *LWF/LI values and Likert scores. Kruskal-Wallis test of Likert vs. 32-echo LWF (left) and 8-echo LI (right). For LWF, median value for Likert 2,3,4 and 5 are: 0.379, 0.182, 0.128 and 0.0464; and median for LI are: 0.231, 0.110, 0.063 and 0.024 respectively.*

Figure 7 illustrates LWF/LI values across biopsy outcomes. LWF/LI values were significantly lower for 'significant' cancer outcomes (Gleason 3+4 and >Gleason 3+4) vs. NS outcomes (negative and Gleason 3+3) (p<0.001). No significant difference was found for LWF/LI within individual outcome subgroups.

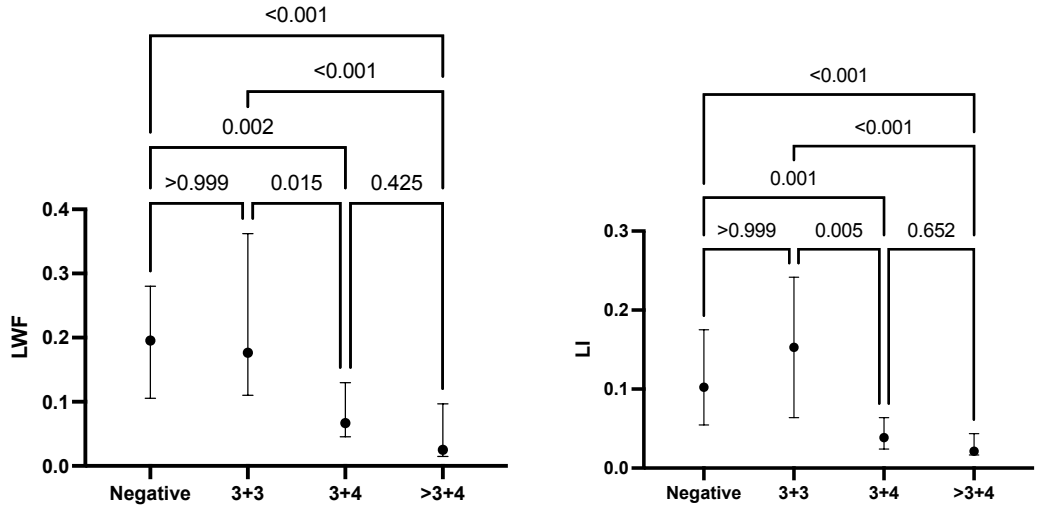


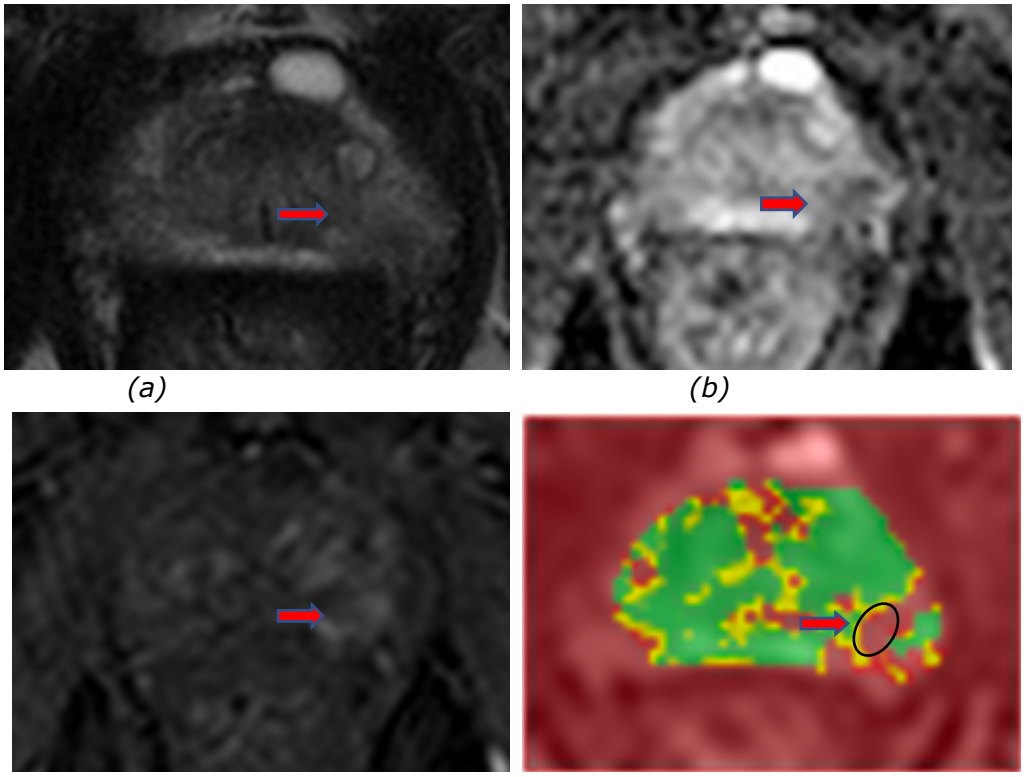
Figure 7. *LWF/LI values and Gleason score. Kruskal-Wallis test comparing of 32-echo LWF (left) and 8-echo LI (right) across individual GL scores (n=109).*

Figure 8 illustrates an example of a mpMRI (a-c) Likert 5 lesion and the corresponding LI map (d); Gleason 3+4 was found at biopsy.

Stage 2: Repeatability and proof-of-concept of full prostate coverage 8-echo LWI

Bland-Altman analysis shows that, for full prostate coverage 8-echo LWI, 95% limits of agreement are -68% to +62%, with a bias of -3%.

There were 19 cases in total that underwent biopsy, one was excluded as biopsy and scan dates were >6 months apart. 18/19 were included for proof-of-concept analysis. 7, 7 and 4 were scored as Likert 3, 4 and 5, respectively. 9 participants had significant cancer (8 with Gleason 3+4, 1 with Gleason > 3+4) and 9 NS biopsy (all negative for cancer). LI classification of biopsied lesions yielded a 89% sensitivity and 78% specificity.



(a) *(b)*

(c) *(d)*

Figure 8. *mpMRI and LI-map. Axial T2 image (a), ADC map (b), DCE image (c) and LI map superimposed on matching luminal water scan for one patient (d). Red arrows show the location of cancer and in (d) the black circle is the lesion ROI.*

Discussion

Our study provided evidence that, despite the difference in absolute values, the 8-echo derived LI provided the same diagnostic performance as 32-echo derived LWF.

We also demonstrated that LI values decrease as Likert/PIRADS scores increase and that significant differences of LI exist between biopsied regions harbouring significant cancer (≥Gleason 3+4) versus no cancer or insignificant disease. In stage 2 of the study, 8-echo LWI was performed with full prostate coverage in a temporally separated cohort, demonstrating good repeatability of LI values and promising sensitivity/specificity for clinically significant cancer.

One of the current clinical challenges in the diagnostic workup of prostate cancer is to improve the specificity of mpMRI. In a recent review, the positive predictive value for clinically significant prostate cancer were 13%, 40%, and 69% for, respectively, PI-RADS 3, 4, and 5 (Mazzone *et al*, 2021); this means that almost half of men currently underwent unnecessary biopsies. Our result on the application of a LI threshold of 0.095 using a full prostate coverage 8-echo sequence showed a putative sensitivity of 89% and specificity at 78% for characterising lesions where decision to biopsy has been made on mpMRI. These findings are encouraging and suggest that the addition of LI-MRI to a mpMRI protocol could potentially result in a 67% reduction of unnecessary biopsies at the expense of 11% (1/9) of men with significant cancer having a potential delay in biopsy.

To validate any new potential biomarker, it is important to compare it with others that are already available. Specifically, the reported summary AUC for ADC is 67% (Shaish *et al*, 2017) for classification of significant prostate cancer. In relation to this, LI generated from 8-echo sequence has shown a potentially better classification performance with an AUC of 0.89. The summarized sensitivity and specificity of ADC values from a recent meta-analysis for separating high-risk from low-risk PCa were 76.9% and 77.0% respectively. The summary ROC curve showed that at sensitivity of 90%, ADC shows a specificity of 60% (Shaish *et al*, 2017). In our study, at 90% sensitivity LI has a specificity of 70%. It should also be noted that unlike diffusion weighted imaging, T2 weighted imaging suffers less from susceptibility artifacts and therefore may have more robustness in the clinical setting (Kozlowski *et al*, 2008).

For any prostate imaging biomarker to be useful it should both be repeatable and produce different values in healthy gland versus cancer. The smaller the variation in quantitative value on repeat measures, and the greater the difference in value between disease and healthy tissue, the more likely it is that the biomarker will show good classification performance. For example, the reported ADC values of clinically

significant prostate cancer (GS≥7) is 0.86 × 10− 3 mm2/s [95% CI 0.83–0.90] and for insignificant prostate cancer (GS 5 and 6) was 1.1 × 10− 3 mm2/s [95% CI 1.03–1.18] (Meyer *et al*, 2020). The difference between significant and insignificant disease is therefore approximately 30% (percentage is calculated as the absolute difference between two means divided by the smaller mean value), and the reported back-to-back repeatability for ADC is relatively small, in between 10-20% (Fedorov *et al*, 2017). Of note, our study showed that LI value for clinically significant prostate cancer was 0.03 [95% CI 0.03-0.05], while for non-significant cancer (benign biopsy or GL3+3) was 0.12 [95% CI 0.11 – 0.15], meaning a difference larger than 300%. The repeatability measurement for LI was moderate at ~70%. Compared to ADC which has 30% difference between significant and insignificant cancer and 10-20% repeatability which is 33-67% of the difference, LI has a difference of 300% and a 70% repeatability which is 23% of the difference; These data suggest at least a similar repeatability value of LI compared to ADC.

LWI may also have the potential to provide a one sequence rapid MRI study for disease detection. A rapid MRI protocol could have significant health economic benefits as well as enabling MRI as a primary test for screening of prostate cancer.

Our study has several limitations. First, we did not test the performance of LI in a setting where it is used by radiologists to score MRI images. Second, a limited number of transition zone tumours were included, and the majority of the quantitative data were derived from the peripheral zone. It is likely that the peripheral zone has a different thresholds differentiating tumour and healthy tissue compared to transition zone; larger datasets are needed to address this issue. Finally, all participants were scanned in single centre on a single Phillips 3T scanner; consequently, we can't make assumptions on the repeatability of our findings using different MRI scanners. To make this technique clinically useful, assessing generalisability will be critical and is part of ongoing work.

In conclusion, LWI using a simplified 8-echo protocol is feasible, repeatable and may help increasing the specificity of mpMRI for the detection of clinically significant prostate cancer.

Disclosure

LI-MRI quantitative values were derived by a software created by Ms Fiona Gong at the UCL Centre for Medical Imaging (2nd Floor Charles Bell House, 43-45 Foley Street, W1W 7TS, London, UK), under the supervision of Prof. Shonit Punwani and Prof. David Atkinson. The pipeline for Luminal Water Imaging acquisition and processing is patented by UCL and is protected by IP regulations. My role in the study

was confirmation of location and scoring of lesions, lesion contouring, histological correlation, paper draft preparation and revision. The manuscript is currently under final internal revision for submission.

Clinical application of LI-MRI

Our preliminary data showed that a shortened LI-MRI protocol (8-echo sequence) is feasible and that the quantitative analysis of LI values may improve the classification of PCa.

However, a further development of the technique was required to make it suitable for implementation in routine clinical practice. Specifically, LI-maps that are generated through ME-T2 image processing were used to extract quantitative LI-values but were not suitable for a qualitative evaluation by radiologists (Figure 9).

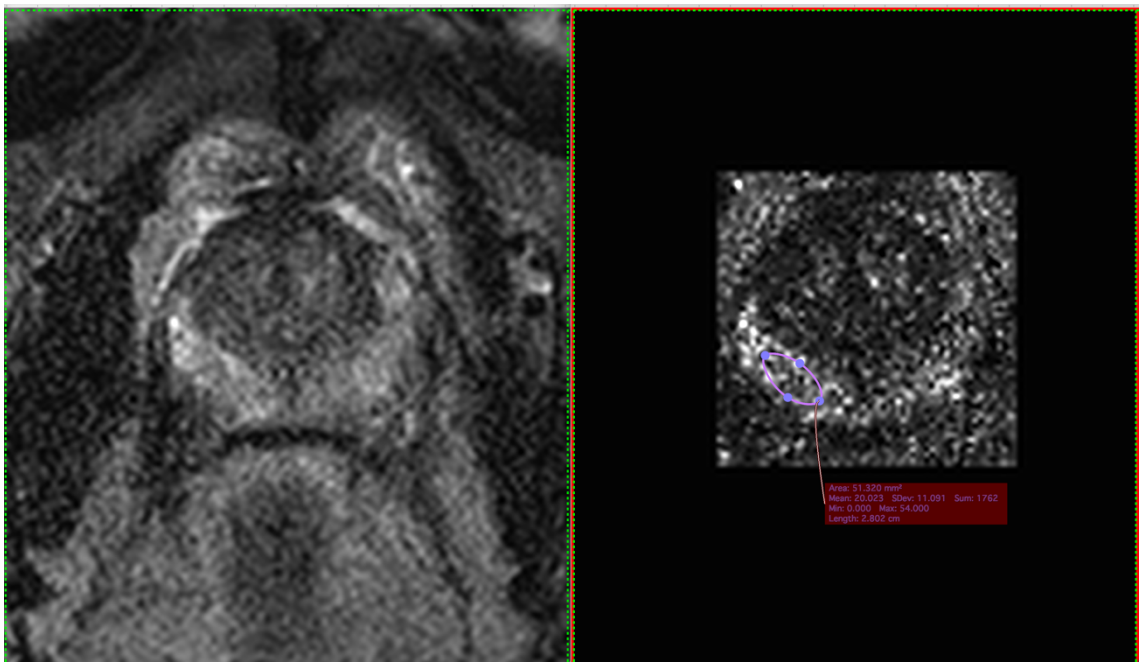


Figure 9. *ME-T2 sequences and LI-maps. On the right the 3rd echo of the ME-T2 sequence; on the left the raw (grayscale) LI-MAP with ROI for LI-value extraction.*

Hence, our goals were:

- to develop a clinically usable LI-MRI layout to allow the qualitative evaluation of LI-map (along with quantitative evaluation);
- to define the criteria for scoring LI-MRI images (ME-T2 and LI-maps);
- to create a dedicated scoring system for LI-MRI

Development of a LI-MRI layout for qualitative evaluation

The main goal of this step was to include the quantitative information of the raw LI-maps into a visual map that allows the qualitative evaluation of the presence/absence/location of lesions, in a similar fashion to the ADC map. Also, we focused on the automatization of the process using a dedicated software and on the

implementation of this tool into an existing clinical PACS suite for visualization and interpretation by radiologists.

Creation of LI-MRI colormaps

Our approach consisted in creating a colormap that displays 3 different colours according to the LI-value of each pixel, ideally: red for pathological LI-values; green for normal LI-values; yellow for indetermined LI-values. The thresholds of LI-values used to generate the different colours were derived from our preliminary data (Figure 10).

The pipeline from ME-T2 acquisition and processing is patented by UCL and is protected under intellectual property regulations.

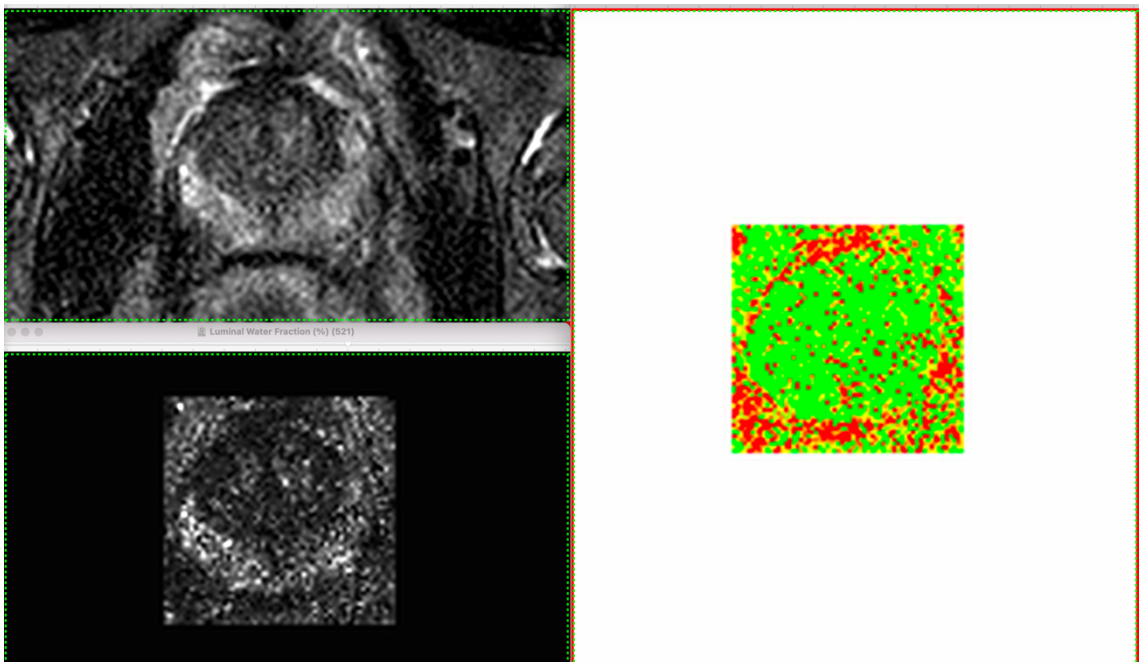


Figure 10. *Greyscale & colormap of LI values. Top left: ME-T2; bottom left: raw grayscale LI-map; right: LI-colormap*

Despite the successful generation of diagnostic colormaps, the process still had several issues to address.

Manual VS automatic contouring & processing

To process ME-T2 images into LI-maps, prostate contouring is required. The contouring process is time-consuming and requires a dedicated radiologist to be done. Also, LI-map processing required high expertise with the MATLAB tools. Hence, a dedicated person was needed to import images/contours into MATLAB and manually start the processing job.

These steps represented an obvious issue that could prevent the application of the LI-MRI technique into clinical practice. Therefore, we focused on the optimization and automatization of the process. This task was accomplished thanks to the collaboration with dedicated engineers of the iCad Inc. company (98 Spit Brook Road, Suite 100 Nashua, NH 03062 USA), a provider of AI-based medical technologies with focus on imaging.

1- The first step was to refine a tool for automatic prostate segmentation, provided by iCad and called Pseg. This step was accomplished in two ways:
    a. Initial refinement: 500 fully anonymized small FOV axial T2-weighted series, derived from a single centre in the USA, were contoured (PZ and TZ contours) to refine the software ability to recognise the prostate boundaries from base to apex.
    b. Fine tuning on low-resolution ME-T2 images. Prostate contours were obtained in 150 patients from the Reimagine trial, drawn on the 3rd echo of the ME-T2 sequences. This allowed the training of the software to correctly recognize the prostate contours on ME-T2 sequences that have lower spatial/contrast resolution compared to the clinical T2W sequences of mpMRI protocols.
2- The second step was to integrate the contouring software output with the processing algorithm to allow a single tool for one-step LI-map generation. This tool was provided by iCad using the refined Pseg software and the Matlab script for LI-map generation as created at UCL.

Implementation of LI-MRI on a clinical PACS suite

To allow radiologist to visualize, interpret and score LI-MRI, we worked on the implementation of the iCad LI-MRI tool into a clinical PACS suite (Biotronics 3Dnet; Biotronics3D, 5 Greenwich View Place, London E149NN). This suite is used widely across UK as a cloud based PACS system both for research and clinical activities. This step was fundamental not only to allow future clinical application of LI-MRI in research trials, but also to pave the way to future implementation into other PACS systems.

To accomplish this task, the iCad tool was installed on the Biotronics3D cloud server. Biotronics3D was programmed to automatically recognise LI-MRI sequences and to launch automatic processing once any ME-T2 sequence is uploaded on their platform.

Finally, we refined the online PACS-RIS suite for LI-MRI visualization, interpretation and scoring (Figure 11).

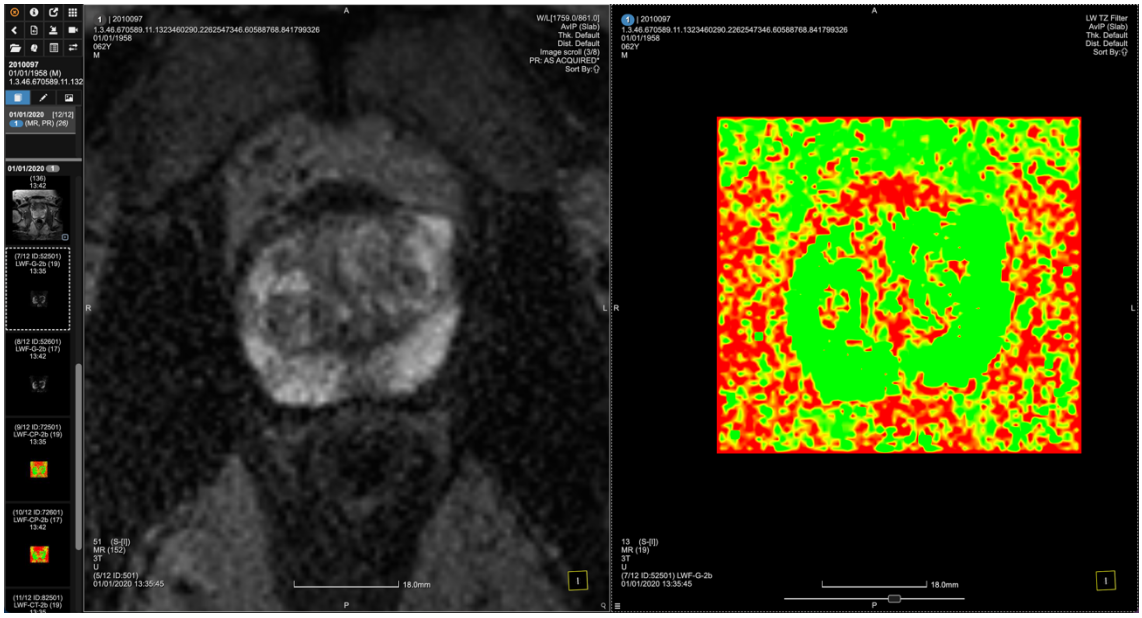


Figure 11. *LI-MRI implementation on Biotronics3D suite.*

Criteria for LI-MRI interpretation & scoring

The criteria for LI-MRI were developed using the INNOVATE and Reimagine trial cohorts, in which patients underwent clinical MRI plus research LI-MRI sequences. We retrospectively evaluated all patients in these cohorts to identify patterns of LI-MRI in normal tissue and in prostatic lesions.

To exploit as many information as possible from the LI-MRI technique, LI-maps should not be evaluated alone but in conjunction with the ME-T2 sequences.

Multiecho-T2 images interpretation

Since multiecho images are basically low resolution T2-weighted images, the PI-RADS criteria/lexicon for T2W can be utilized for interpretation and scoring of ME-T2 (Turkbey *et al*, 2019).

Caveat for TZ lesions. Since the spatial resolution of ME-T2 is lower compared to standard SFOV axial T2W images of a clinical mpMRI protocol, the typical finding of blurred margins (also called "erased charcoal sign") may not be clearly visible. On ME-T2, TZ lesions may appear as areas of lower SI compared to the surrounding TZ, but may have rather delimited margins (Figure 12).

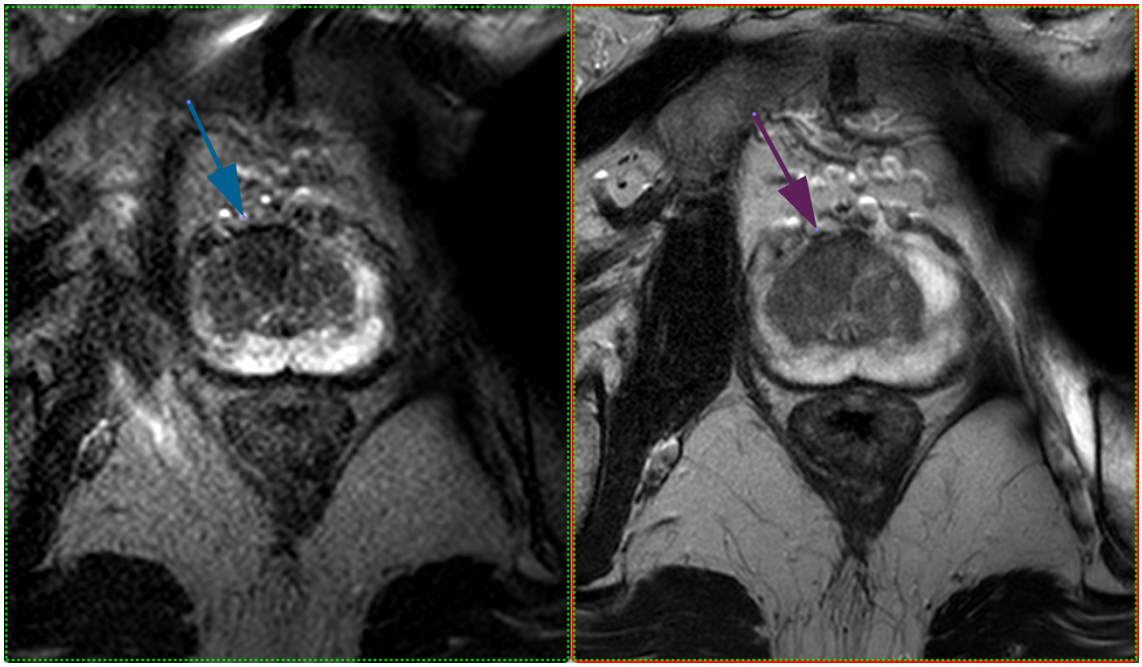


Figure 12. *Right anterior TZ lesion. Right image (axial T2W): right anterior TZ lesion (PI-RADS 5) with homogeneous low T2 SI and blurred margins. Left image (ME-T2): the same lesion appears as a delimited area with lower SI compared to the contralateral (normal) TZ, while the blurred margins are not clearly visible.*

LI-MRI colormap interpretation

Normal prostate tissue should have homogeneous green signal both in the PZ and the TZ. A common alternative pattern of normal tissue can be the loosely scattered pattern (yellow or red dots on a green background) (Figure 13). The scattered pattern is almost invariably present in the TZ given its heterogeneity but can be found also in the PZ.

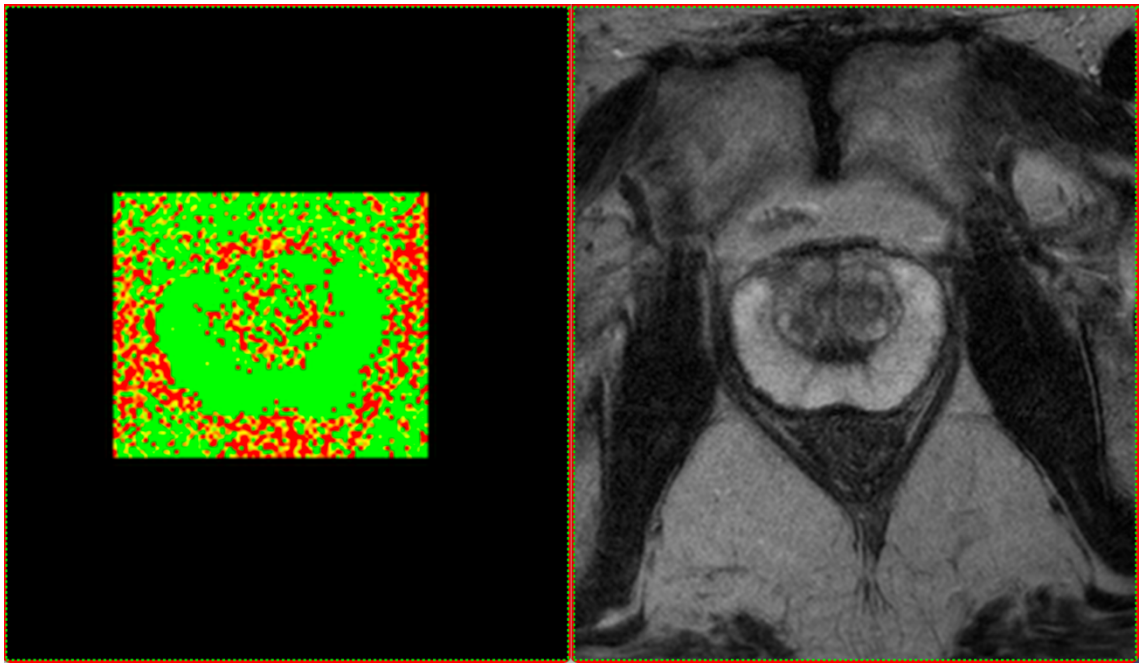


Figure 13. *Normal prostate on LI-map. LI-map (left image) showing homogeneous green PZ and scattered pattern in the TZ.*

Suspicious lesions are represented by "clusters" of red pixels, i.e. focal lesion with predominantly red (and yellow) components (Figure 14).

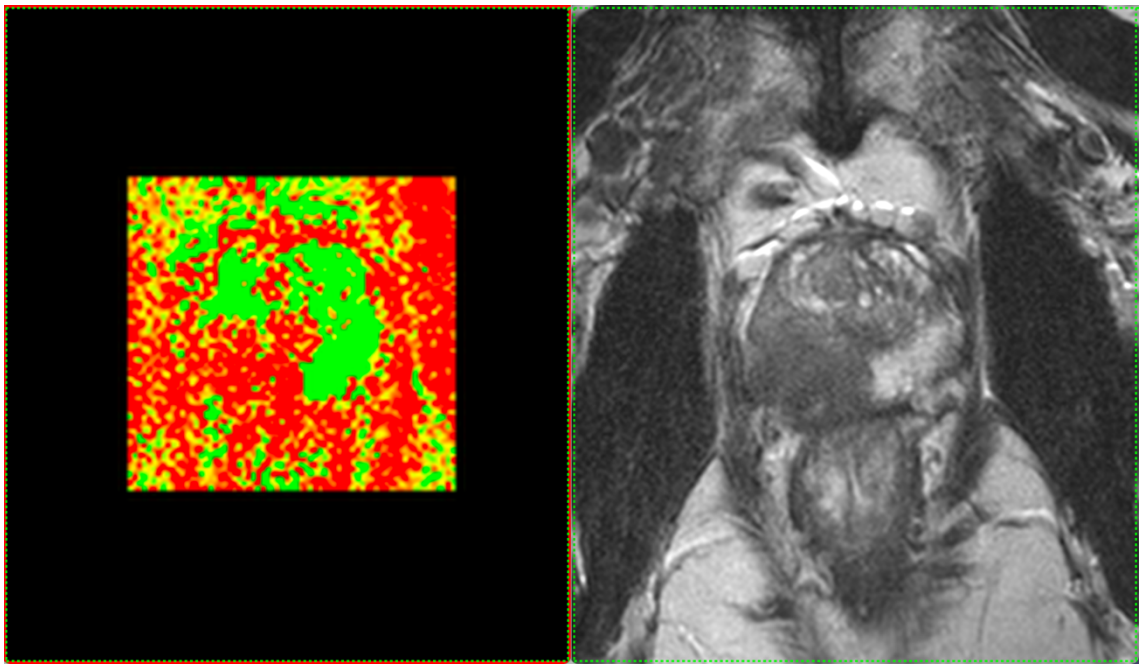


Figure 14. *PI-RADS 5 PZ lesion. Right PZ lesion (PI-RADS 5). On LI-map (left image) there is obvious focal red signal corresponding to the PZ lesion on T2W (right image).*

Findings that cannot be interpreted as unequivocally benign or unequivocally pathological should be considered indetermined on LI-maps. One of the more

frequent indetermined pattern is the densely scattered, or "geographical" pattern, where an area of the prostate has inhomogeneous altered signal (mix of red-yelow-green pixels) (Figure 15).

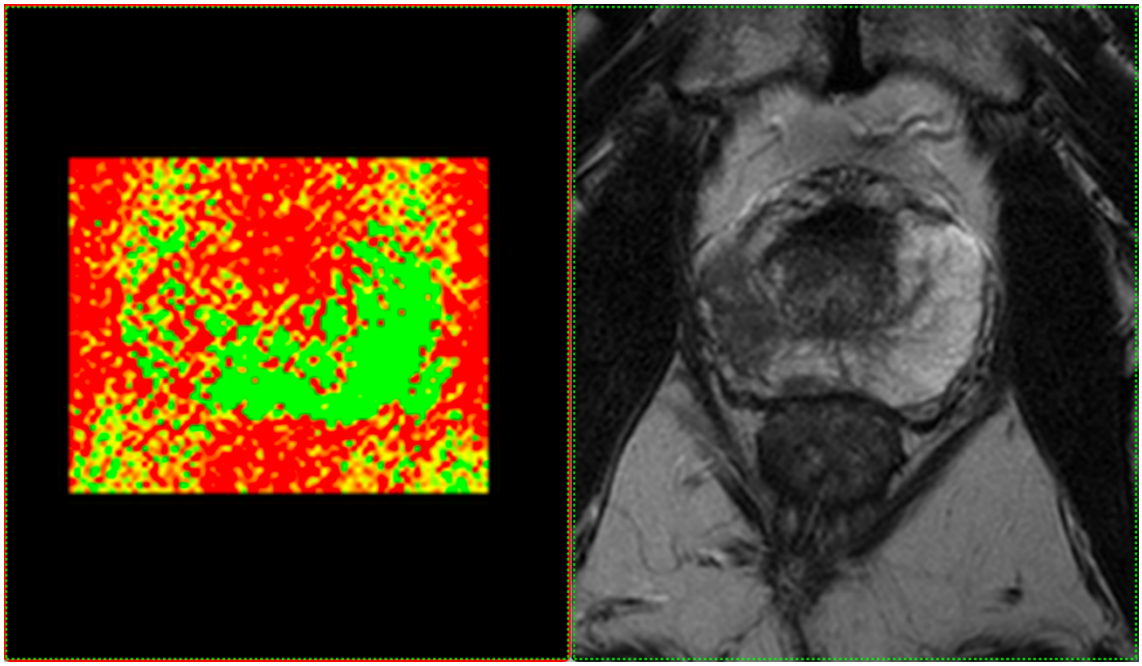


Figure 15. *Indetermined findings on LI-map. Inhomogeneous SI of the right PZ on LI-map (left image) with interspersed (predominantly) red-yellow areas on a green background. The lesion was judged equivocal (PI-RADS 3) also on biparametric MRI (right image: axial T2W).*

Caveats for LI-maps. Normal structures with fibrous components appear as red areas on the map but should not be mistaken for PCa. For example, the fibrous scaffold of the TZ (typicaly at the PZ/TZ interface) and the AFMS appear red due to very low SI on T2 (FigureX). A correlation with morphological sequences (T2W) is necessary to avoid this pitfall.

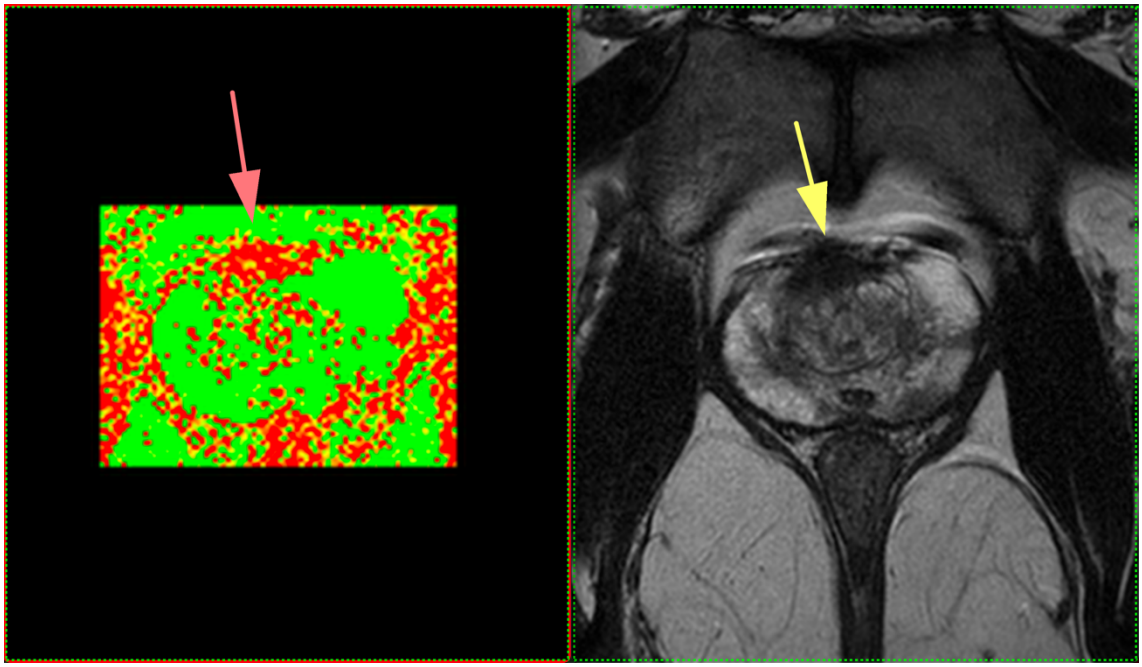


Figure 16. *Anterior fibromuscular stroma (AFMS). Prominent AFMS as seen on LI-maps (left) and T2W.*

LI-MRI scoring system (LUMINAL-RADS)

Based on the above observations, we proposed a scoring system to interpret LI-MRI images. The final score depends on the findings of both the ME-T2 and the LI-MRI images. Similarly to PI-RADS, we decide to include a dominant sequence, in this case the ME-T2 sequence. We believe that, at least at this early stage, it is safer to rely on a better-known imaging pattern (i.e., T2-weighted imaging) rather than a totally new one (LI-maps) to determine the presence/absence of PCa. Also, this should allow for a better inter-reader agreement. With further refinements of the technique, and after reviewing imaging-pathological correspondence on future studies, the scoring system is likely destined to be refined.

Multi-echo T2 score.

PZ:

1- Uniform hyperintense signal intensity (normal)
2- Linear or wedge-shaped hypointensity or diffuse mild hypointensity, usually indistinct margin
3- Heterogeneous signal intensity or non-circumscribed, rounded, moderate hypointensity. Includes others that do not qualify as 2, 4, or 5
4- Circumscribed, homogenous hypointense focus/mass <1.5cm in greatest dimension
5- Same as 4 but ≥1.5cm in greatest dimension or evidence of ECE

TZ:

1- Hyperintense areas of the TZ
2- Isointense areas of TZ, or mildly hypointense clearly circumscribed and/or capsulated (typical nodule); thin homogeneous hypointense area between nodules (fibrous stroma).
3- Others that do not qualify as 2, 4, or 5
4- Lenticular or non-circumscribed, homogeneous, hypointense; causing profile bulge; loss of normal nodular texture of TZ and replacement by homogeneous reduced T2 signal
5- Same as 4, but ≥1.5cm in greatest dimension or evidence of ECE

AFMS:

1- No evidence of AFMS
2- Thin hypointense band in the anatomic region of the AFMS
3- Thick band of low signal in the anatomic location of the AFMS that symmetrically alters the contour of the prostate
4- Thick asymmetric band of low signal in the anatomic location of the AFMS, causing bulge or alteration of the prostatic profile
5- Same as 4 but ≥1.5cm in greatest dimension or evidence of ECE

LI-MRI colormap score.

PZ:

1- Normal appearing map (i.e.: homogeneously green)
2- Noise-like red on green background; linear or wedge-shaped
3- Others that do not qualify as 1-2 or 4-5
4- Homogeneous, red focal mass; <1.5cm in greatest dimension
5- Same as 4 but >1.5cm in greatest dimension

TZ:

1- Normal appearing map (i.e.: homogeneously green)
2- Noise-like red on green background; linear or curvilinear red on green background
3- Others that do not qualify as 1-2 or 4-5
4- Homogeneous, non-nodular red focal mass; <1.5cm in greatest dimension
5- Same as 4 but >1.5cm in greatest dimension

AFMS:

0- No red curvilinear mass at site of AFMS

1- Red curvilinear mass at site of AFMS

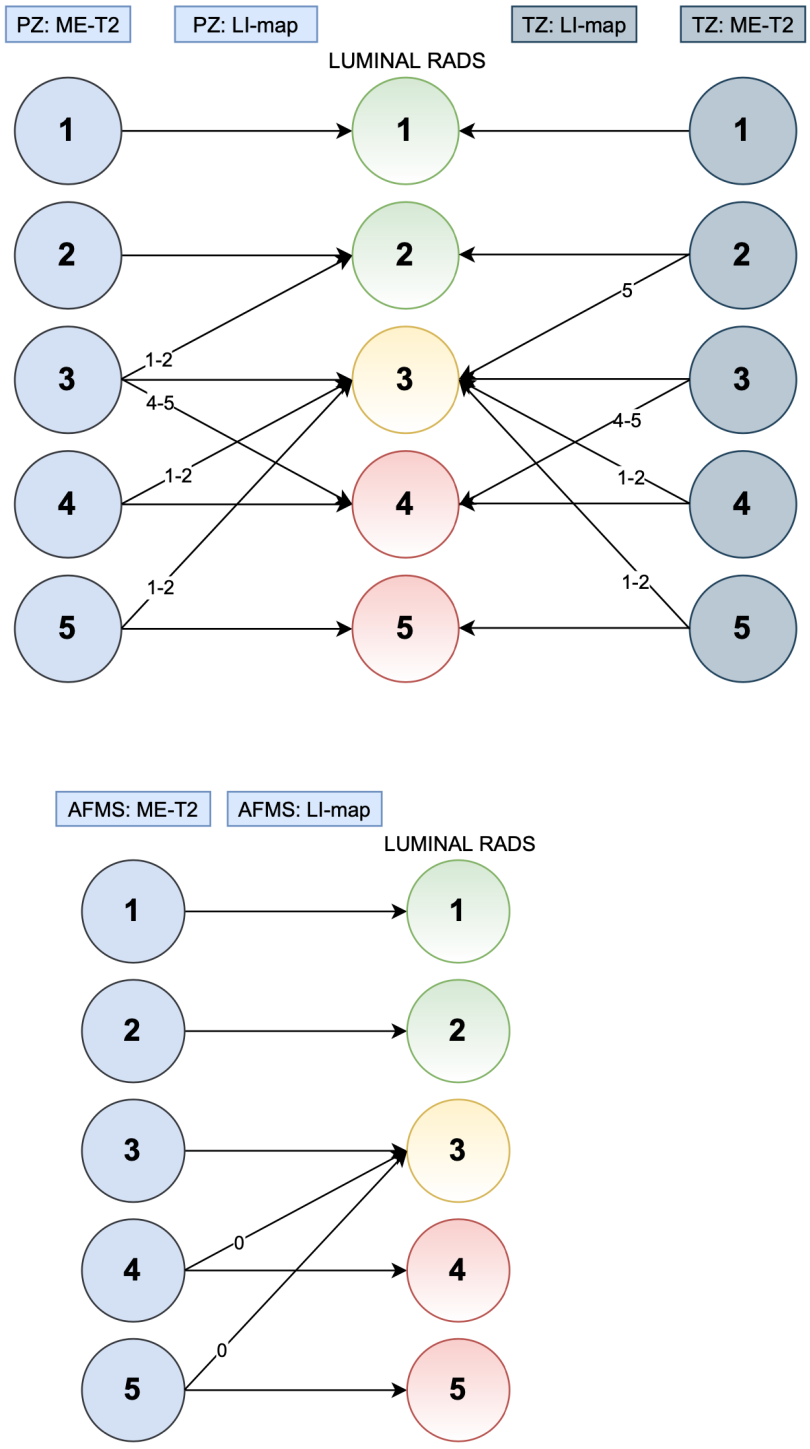


Figure 17. *LUMINAL-RADS scheme.*

Discussion

Multiparametric MRI has become the imaging modality of choice for the diagnostic workup of PCa (Mottet *et al*, 2020). MpMRI detects csPCa with high sensitivity (>90%) with a high NPV (approximately 90%) in a routine diagnostic setting (Ahmed *et al*, 2017). However, despite its widespread use in clinical practice, there are still several limitations that need to be addressed.

The reproducibility of prostate MRI interpretation across different readers has traditionally been considered a major weakness of MRI. In fact, one of the main aims of the PI-RADS scoring system was to standardize prostate MRI reporting. Still, a number of studies reported unsatisfactory interreader agreement of MRI (Park *et al*, 2020). However, most of the studies showed significant flaws in the study design and in the statistical analyses used. Thus, there was an urgent need for clarification of the actual reliability of MRI interpretation in the real-world setting. To address this issue, we conducted a multireader study designed to replicate the routine clinical practice, using robust statistical methods, and we observed that the reproducibility of PI-RADS scores is actually better that previously reported (Brembilla *et al*, 2020a). The importance of our findings should not be overlooked. To be clinically useful, any diagnostic tool should be not only accurate but also reproducible. Thus, it is fundamental to have a realistic insight on both the accuracy and reproducibility of prostate MRI. Also, this allows a reliable and critical evaluation of any proposed refinement/evolution of the MRI technique (e.g., the introduction of novel sequences or the update of scoring-systems). Thus, we believe that our findings could represent a milestone in the evaluation of MRI reproducibility in a real-world setting using PI-RADS and can be used as a standard of reference when evaluating new developments of MRI.

Accessibility of MRI is another major issue in the PCa diagnostic pathway. MpMRI is a costly examination with long scan/procedural times. Consequently, the widespread application of mpMRI is facing a limited availability of dedicated scanners and personnel, resulting in long waiting times. The need for simplification of the MRI diagnostic pathway is becoming clear and the use of non-contrast MRI protocols (i.e., bi-parametric MRI) has been proposed to address this issue. In this context, bi-parametric MRI offers several potential advantages over mpMRI. Among the others, non-contrast protocols may reduce scan/procedural times (e.g., pre-scan consents, cannulation, post-scan observation), increasing the MRI scan slots available per day. Along with "full" biparametric MRI protocols (i.e., multiplanar T2w and DWI-ADC), also abbreviated or "fast" bpMRI protocols have been investigated that generally

consist in monoplanar T2W + DWI (+/- ADC maps) and require approximately 10 minutes to be performed (van der Leest *et al*, 2019b). Since a-bpMRI is associated with an extreme reduction of scan times, it could have important health-economic implications. Also, the availability of such ultra-fast protocols paved the way to research trials that investigated the potential use of a-bpMRI as a tool for population-based screening (Marsden *et al*, 2021a; Eldred-Evans *et al*, 2021). However, despite evidence is accumulating on the diagnostic performance of full bpMRI protocols (Bass *et al*, 2020), only few studies to date specifically focused on a-bpMRI (van der Leest *et al*, 2019b), and high-quality evidence on their performance is lacking. To fill this gap, we aimed to investigate the performance of a-bpMRI and compare it to bpMRI and mpMRI using a robust study design (Brembilla *et al*, 2022). We performed a multireader evaluation split in two separate sessions using TTPM biopsies as the standard of reference. Of note, biopsies were performed in all patients regardless the MRI results. Our findings confirmed the hypothesis that there is no significant difference in the diagnostic yield of a-bpMRI and bpMRI/mpMRI in expert hands. Moreover, we simulated the application of a-bpMRI to a low prevalence setting (e.g. screening), and we observed that higher positivity cut-off should be used in those populations without compromising the NPV of MRI (given the very low prevalence). Despite we did not find an increased rate of indetermined findings with a-bpMRI (MRI score = 3), as previously described (van der Leest *et al*, 2019b), we observed that the interreader agreement of short protocols is lower than full protocols. In all, we provided evidence that ultra-fast protocols are feasible and may have a good performance if interpreted by expert readers; to address the presumed increased variability, at least in a screening setting, a double reading approach can be considered. At the same time, we acknowledge that further evidence is needed before we can implement bpMRI in the clinical practice. Also, scan quality is even more important for bpMRI than for mpMRI, since there is no DCE as a safety net when artefact affect the quality of DWI.

Despite its high sensitivity, MRI has low specificity for csPCa (Drost *et al*, 2019) and the rate of false positive findings is high (low PPV) especially for lower MRI scores (3-4). In general terms, MRI has high contrast resolution but lacks biologic specificity. Attempts have been made to address this issue by including clinical data into the decisional algorithm for biopsy. Among the others, PSA density has shown promising results in reducing the rate of unnecessary biopsies for equivocal MRI lesions (i.e., MRI score = 3) (Falagario *et al*, 2021). Our project aimed at investigating a novel MRI technique called luminal index MRI (LI-MRI) as a potential tool for better characterization of PCa. The biological rationale lies in the observation, based on

histology studies, that csPCa alters the balance of glandular to cellular spaces within the prostate glands, reducing the fraction of glandular lumen (i.e., the luminal water fraction); this fraction decreases further as the Gleason grade of tumour increases. Accordingly, LI-MRI allows the modelling of two compartments within the prostate gland, the luminal space (long T2 values) and the stromal + epithelial spaces (short T2 values) (Devine et al, 2019), allowing a voxel-based estimation of the luminal water fraction. Hence, LI-MRI has the potential to reflect prostate tissue composition better than conventional T2W sequences and DWI. The original technique employed non-standard MRI sequences (64-echoes ME-T2) which required complex set-up and provided one image slice taken through the whole prostate (Sabouri *et al*, 2017). Subsequently, Devine et al. showed the feasibility of a simplified LI-MRI protocol (32-echoes ME-T2) sampling the whole prostate (Devine et al, 2019); however, low spatial resolution and long scan times still limited its applicability. In our project, we took another step forward by investigating the performance of an abbreviated LI-MRI protocol (8-echoes ME-T2; less that 10 min scan with whole prostate coverage and higher spatial resolution), its reproducibility and its potential application to the clinical practice. Our result showed that that LI values correlate with the presence of csPCa, yielding a similar sensitivity (90%) but higher specificity (70% vs 50%) than mpMRI in the detection of csPCa. In all, we provided evidence that 8-echo ME-T2 LI-MRI is feasible, reproducible, and that it can help mpMRI increasing its specificity for the characterization of csPCa. Also, 8-echoes LI-MRI can be performed in less than 10 minutes and is not prone to the typical artefacts that degrade the quality of DWI images.

Of note, our study was based on a quantitative evaluation of LI-values in pre-determined lesions (reported at mpMRI), thus its performance as a tool to detect csPCa is not known. A further evolution of the technique was required to make it suitable for qualitative evaluation in a clinical setting. Specifically, the raw LI-maps that are generated by processing ME-T2 images were originally intended to provide a quantitative estimation of the LI-values and are not suitable for lesion detection by radiologists. Furthermore, manual contouring of the prostate was required for LI-maps processing. Finally, processing was a time-consuming process that was possible only using a dedicated MATLAB algorithm. We therefore worked at the development of an automated tool for segmentation and processing, to generate both a quantitative map and a colormap that was suitable for lesion detection. This was made possible by the collaboration with an external company called iCAD (processing) that developed an AI-based software for automatic segmentation and subsequent processing that generates both raw and colormaps. The process was

tested by implementing the software into a cloud based PACS/RIS system (Biotronics3D). This platform will be used for LI-MRI training and future clinical/research applications.

Finally, based on our retrospective analyses of patients who underwent LI-MRI and biopsy, we proposed the first scoring system for LI-MRI. It is based on the evaluation of both the ME-T2 sequences and the LI-MRI maps to generate a final score from 1 (very low likelihood of PCa) to 5 (very high likelihood of PCa) conceptually similar to PI-RADS. Future validation is needed to test the performance of LI-MRI alone or in combination with mpMRI, and the performance of its scoring system.

To summarize, we provided evidence that a simplified, ultra-fast MRI protocol may perform as well as mpMRI in expert hands, addressing some major limitations of MRI and specifically its limited accessibility. Also, it could represent the ideal tool to explore the feasibility of population-based screening programs using MRI. Also, we demonstrated the feasibility, applicability and reproducibility of LI-MRI; given its high specificity, it could help refining the characterization of csPCa of mpMRI. Potentially, LI-MRI could also represent another potential tool for population-based screening considering short scan times (<10 min), absence of contrast media and reduced rate of susceptibility artefacts compared to DWI.

Future applications

The results of our project paved the way to two prospective, multicentre trials that will be carried out in the UK.

The CLIMATE study (Comparison of diagnostic accuracy of Luminal Index and Multi-parametric MRI for accelerated detection of significant prostate cancer) is a prospective, multi-centre paired, non-randomised, comparative study of LI-MRI and mp-MRI targeted biopsy in participants with suspected of prostate cancer. It will recruit 702 men over 3 years across the UK and will validate the clinical applicability and performance of LI-MRI when used alone or in adjunction of mpMRI.

The LIMIT study (Luminal Index MRI Identification of Treatment critical Prostate Cancer) will invite men for prostate cancer screening using luminal index, bi-parametric MRI and PSA density, to identify men with suspicious lesions and to evaluate acceptability.

References

Ahmed HU, El-Shater Bosaily A, Brown LC, Gabe R, Kaplan R, Parmar MK, Collaco-Moraes Y, Ward K, Hindley RG, Freeman A, *et al* (2017) Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet*

Ahmed HU, Hu Y, Carter T, Arumainayagam N, Lecornet E, Freeman A, Hawkes D, Barratt DC & Emberton M (2011) Characterizing clinically significant prostate cancer using template prostate mapping biopsy. *J Urol* 186: 458–464

Altman DG & Bland j. M (1994) Statistics Notes: Diagnostic tests 2: Predictive values. *BMJ* 309: 102

Barentsz JO, Richenberg J, Clements R, Choyke P, Verma S, Villeirs G, Rouviere O, Logager V & Fütterer JJ (2012) ESUR prostate MR guidelines 2012. *Eur Radiol*

Barth BK, De Visschere PJL, Cornelius A, Nicolau C, Vargas HA, Eberli D & Donati OF (2017) Detection of clinically significant prostate cancer: Short dual-pulse sequence versus standard multiparametric MR Imaging - A multireader study. *Radiology* 284: 725–736

Bass EJ, Pantovic A, Connor M, Gabe R, Padhani AR, Rockall A, Sokhi H, Tam H, Winkler M & Ahmed HU (2020) A systematic review and meta-analysis of the diagnostic accuracy of biparametric prostate MRI for prostate cancer in men at risk. *Prostate Cancer Prostatic Dis* doi:10.1038/s41391-020-00298-w [PREPRINT]

Bell KJL, Del Mar C, Wright G, Dickinson J & Glasziou P (2015) Prevalence of incidental prostate cancer: A systematic review of autopsy studies. *Int J Cancer* 137: 1749–1757

Bill-Axelson A, Holmberg L, Garmo H, Taari K, Busch C, Nordling S, Häggman M, Andersson S-O, Andrén O, Steineck G, *et al* (2018) Radical Prostatectomy or Watchful Waiting in Prostate Cancer - 29-Year Follow-up. *N Engl J Med* 379: 2319–2329

Bosaily AES, Frangou E, Ahmed HU, Emberton M, Punwani S, Kaplan R, Brown LC, Freeman A, Jameson C, Hindley R, *et al* (2020) Additional Value of Dynamic Contrast-enhanced Sequences in Multiparametric Prostate Magnetic Resonance Imaging: Data from the PROMIS Study[Formula presented]. *Eur Urol* 78: 503–511

Brembilla G, Dell'Oglio P, Stabile A, Damascelli A, Brunetti L, Ravelli S, Cristel G, Schiani E, Venturini E, Grippaldi D, *et al* (2020a) Interreader variability in prostate MRI reporting using Prostate Imaging Reporting and Data System

version 2.1. *Eur Radiol*

Brembilla G, Giganti F, Sidhu H, Imbriaco M, Mallett S, Stabile A, Freeman A, Ahmed HU, Moore C, Emberton M, *et al* (2022) Diagnostic Accuracy of Abbreviated Bi-Parametric MRI (a-bpMRI) for Prostate Cancer Detection and Screening: A Multi-Reader Study. *Diagnostics* 12: 231

Brembilla G, Takwoingi Y & Kasivisvanathan V (2020b) Tackling Interobserver Variability in Multiparametric Magnetic Resonance Imaging (MRI): Is MRI Even Better than We Think for Prostate Cancer Diagnosis? *Eur Urol*: 43–45

Brizmohun Appayya M, Adshead J, Ahmed HU, Allen C, Bainbridge A, Barrett T, Giganti F, Graham J, Haslam P, Johnston EW, *et al* (2018) National implementation of multi-parametric magnetic resonance imaging for prostate cancer detection – recommendations from a UK consensus meeting. *BJU Int* 122: 13–25 doi:10.1111/bju.14361 [PREPRINT]

Caglic I, Hansen NL, Slough RA, Patterson AJ & Barrett T (2017) Evaluating the effect of rectal distension on prostate multiparametric MRI image quality. *Eur J Radiol* 90: 174–180

Cereser L, Giannarini G, Bonato F, Pizzolitto S, Como G, Valotto C, Ficarra V, Dal Moro F, Zuiani C & Girometti R (2020) Comparison of multiple abbreviated multiparametric MRI-derived protocols for the detection of clinically significant prostate cancer. *Ital J Urol Nephrol*

Cicchetti D V. & Feinstein AR (1990) High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol*

Conger AJ (1980) Integration and generalization of kappas for multiple raters. *Psychol Bull*

Dell'Oglio P, Stabile A, Soligo M, Brembilla G, Esposito A, Gandaglia G, Fossati N, Bravi CA, Dehò F, De Cobelli F, *et al* (2019) There Is No Way to Avoid Systematic Prostate Biopsies in Addition to Multiparametric Magnetic Resonance Imaging Targeted Biopsies. *Eur Urol Oncol*: 1–7

Devine W, Giganti F, Johnston EW, Sidhu HS, Panagiotaki E, Punwani S, Alexander DC & Atkinson D (2019) Simplified Luminal Water Imaging for the Detection of Prostate Cancer From Multiecho T2 MR Images. *J Magn Reson Imaging* 50: 910–917

Drost FJH, Osses DF, Nieboer D, Steyerberg EW, Bangma CH, Roobol MJ & Schoots IG (2019) Prostate MRI, with or without MRI-targeted biopsy, and systematic biopsy for detecting prostate cancer. *Cochrane Database Syst Rev* 2019

Eklund M, Jäderling F, Discacciati A, Bergman M, Annerstedt M, Aly M, Glaessgen A, Carlsson S, Grönberg H & Nordström T (2021) MRI-Targeted or Standard

Biopsy in Prostate Cancer Screening. *N Engl J Med* 385: 908–920

Eldred-Evans D, Burak P, Connor M, Day E, Evans M, Fiorentino F, Gammon M, Hosking-Jervis F, Klimowska- Nassar N, McGuire W, *et al* (2021) Population-based prostate cancer screening with Magnetic Resonance or Ultrasound Imaging: The IP1-PROSTAGRAM study. *JAMA Oncol*

Eldred-Evans D, Tam H, Sokhi H, Padhani AR, Winkler M & Ahmed HU (2020) Rethinking prostate cancer screening: could MRI be an alternative screening test? *Nat Rev Urol* 17: 526–539

Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR & Humphrey PA (2016) 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* 40: 244–252

Falagario UG, Jambor I, Lantz A, Ettala O, Stabile A, Taimen P, Aronen HJ, Knaapila J, Perez IM, Gandaglia G, *et al* (2021) Combined Use of Prostate-specific Antigen Density and Magnetic Resonance Imaging for Prostate Biopsy Decision Planning: A Retrospective Multi-institutional Study Using the Prostate Magnetic Resonance Imaging Outcome Database (PROMOD). *Eur Urol Oncol* 4: 971–979

Fedorov A, Vangel MG, Tempany CM & Fennessy FM (2017) Multiparametric Magnetic Resonance Imaging of the Prostate: Repeatability of Volume and Apparent Diffusion Coefficient Quantification. *Invest Radiol* 52: 538–546

Feinstein AR & Cicchetti D V. (1990) High agreement but low Kappa: I. the problems of two paradoxes. *J Clin Epidemiol*

Gandaglia G, Leni R, Bray F, Fleshner N, Freedland SJ, Kibel A, Stattin P, Van Poppel H & La Vecchia C (2021) Epidemiology and Prevention of Prostate Cancer. *Eur Urol Oncol* 4: 877–892

Garcia-Reyes K, Passoni NM, Palmeri ML, Kauffman CR, Choudhury KR, Polascik TJ & Gupta RT (2015) Detection of prostate cancer with multiparametric MRI (mpMRI): effect of dedicated reader education on accuracy and confidence of index and anterior cancer diagnosis. *Abdom Imaging* 40: 134–142

Gelfond J, Choate K, Ankerst DP, Hernandez J, Leach RJ & Thompson IM (2015) Intermediate-term risk of prostate cancer is directly related to baseline prostate specific antigen: Implications for reducing the burden of prostate specific antigen screening. *J Urol* 194: 46–51

Giganti F, Allen C, Emberton M, Moore CM & Kasivisvanathan V (2020) 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* 0

Girometti R, Giannarini G, Greco F, Isola M, Cereser L, Como G, Sioletic S, Pizzolitto S, Crestani A, Ficarra V, *et al* (2019) Interreader agreement of PI-RADS v. 2 in assessing prostate cancer with multiparametric MRI: A study using whole-mount histology as the standard of reference. *J Magn Reson Imaging* 49: 546–555

Greer MD, Brown AM, Shih JH, Summers RM, Marko J, Law YM, Sankineni S, George AK, Merino MJ, Pinto PA, *et al* (2017) Accuracy and agreement of PIRADSv2 for prostate cancer mpMRI: A multireader study. *J Magn Reson Imaging* 45: 579–585

Greer MD, Shih JH, Lay N, Barrett T, Bittencourt L, Borofsky S, Kabakus I, Law YM, Marko J, Shebel H, *et al* (2019) Interreader variability of prostate imaging reporting and data system version 2 in detecting and assessing prostate cancer lesions at prostate MRI. *Am J Roentgenol* 212: 1197–1205

Grossman DC, Curry SJ, Owens DK, Bibbins-Domingo K, Caughey AB, Davidson KW, Doubeni CA, Ebell M, Epling JW, Kemper AR, *et al* (2018) Screening for prostate cancer USPreventive servicestaskforcerecommendation statement. *JAMA - J Am Med Assoc* 319: 1901–1913

Guichard G, Larré S, Gallina A, Lazar A, Faucon H, Chemama S, Allory Y, Patard JJ, Vordos D, Hoznek A, *et al* (2007) Extended 21-sample needle biopsy protocol for diagnosis of prostate cancer in 1000 consecutive patients. *Eur Urol* 52: 430–435

Guo CC & Epstein JI (2006) Intraductal carcinoma of the prostate on needle biopsy: Histologic features and clinical significance. *Mod Pathol* 19: 1528–1535

Gwet KL (2008) Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol* 61: 29–48

Hötker AM, Dappa E, Mazaheri Y, Ehdaie B, Zheng J, Capanu M, Hricak H & Akin O (2019) The influence of background signal intensity changes on cancer detection in prostate MRI. *Am J Roentgenol* 212: 823–829

Huebner NA, Korn S, Resch I, Grubmüller B, Gross T, Gale R, Kramer G, Poetsch N, Clauser P, Haitel A, *et al* (2020) Visibility of significant prostate cancer on multiparametric magnetic resonance imaging (MRI)—do we still need contrast media? *Eur Radiol*

Hugosson J, Roobol MJ, Månsson M, Tammela TLJ, Zappa M, Nelen V, Kwiatkowski M, Lujan M, Carlsson S V., Talala KM, *et al* (2019) A 16-yr Follow-up of the European Randomized study of Screening for Prostate Cancer(Figure presented.). *Eur Urol* 76: 43–51

Jambor I, Verho J, Ettala O, Knaapila J, Taimen P, Syvänen KT, Kiviniemi A, Kähkönen E, Perez IM, Seppänen M, *et al* (2019) Validation of improd biparametric mri in men with clinically suspected prostate cancer: A prospective multi-institutional trial. *PLoS Med* 16

Kang Z, Min X, Weinreb J, Li Q, Feng Z & Wang L (2019) Abbreviated biparametric versus standard multiparametric MRI for diagnosis of prostate cancer: A systematic review and meta-analysis. *Am J Roentgenol* 212: 357–365

Karanasios E, Caglic I, Zawaideh JP & Barrett T (2022) Prostate MRI quality: clinical impact of the PI-QUAL score in prostate cancer diagnostic work-up. *Br J Radiol*

Kasivisvanathan V, Rannikko AS, Borghi M, Panebianco V, Mynderse LA, Vaarala MH, Briganti A, Budäus L, Hellawell G, Hindley RG, *et al* (2018) MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. *N Engl J Med* 378: 1767–1777

Kozlowski P, Chang SD & Goldenberg SL (2008) Diffusion-weighted MRI in prostate cancer - comparison between single-shot fast spin echo and echo planar imaging sequences. *Magn Reson Imaging* 26: 72–76

Kuhl CK, Bruhn R, Krämer N, Nebelung S, Heidenreich A & Schrading S (2017) Abbreviated biparametric prostate MR imaging in men with elevated prostate-specific antigen. *Radiology* 285: 493–505

Kweldam CF, Wildhagen MF, Steyerberg EW, Bangma CH, Van Der Kwast TH & Van Leenders GJLH (2015) Cribriform growth is highly predictive for postoperative metastasis and disease-specific death in Gleason score 7 prostate cancer. *Mod Pathol* 28: 457–464

Landis JR & Koch GG (1977) The Measurement of Observer Agreement for Categorical Data. *Biometrics*

Latifoltojar A, Appayya MB, Barrett T & Punwani S (2019) Similarities and differences between Likert and PIRADS v2.1 scores of prostate multiparametric MRI: a pictorial review of histology-validated cases. *Clin Radiol* 74: 895.e1-895.e15 doi:10.1016/j.crad.2019.08.020 [PREPRINT]

van der Leest M, Israël B, Cornel EB, Zámecnik P, Schoots IG, van der Lelij H, Padhani AR, Rovers M, van Oort I, Sedelaar M, *et al* (2019a) High Diagnostic Performance of Short Magnetic Resonance Imaging Protocols for Prostate Cancer Detection in Biopsy-naïve Men: The Next Step in Magnetic Resonance Imaging Accessibility. *Eur Urol* 76: 574–581

van der Leest M, Israël B, Cornel EB, Zámecnik P, Schoots IG, van der Lelij H, Padhani AR, Rovers M, van Oort I, Sedelaar M, *et al* (2019b) High Diagnostic Performance of Short Magnetic Resonance Imaging Protocols for Prostate

Cancer Detection in Biopsy-naïve Men: The Next Step in Magnetic Resonance Imaging Accessibility. *Eur Urol* 76: 574–581

Leisenring W, Alonzo T & Pepe MS (2000) Comparisons of predictive values of binary medical diagnostic tests for paired designs. *Biometrics* 56: 345–351

Luzzago S, Petralia G, Musi G, Catellani M, Alessi S, Di Trapani E, Mistretta FA, Serino A, Conti A, Pricolo P, *et al* (2019) Multiparametric Magnetic Resonance Imaging Second Opinion May Reduce the Number of Unnecessary Prostate Biopsies: Time to Improve Radiologists' Training Program? *Clin Genitourin Cancer* 17: 88–96

Marsden T, Lomas DJ, McCartan N, Hadley J, Tuck S, Brown L, Haire A, Moss CL, Green S, Van Hemelrijck M, *et al* (2021a) ReIMAGINE Prostate Cancer Screening Study: protocol for a single-centre feasibility study inviting men for prostate cancer screening using MRI. *BMJ Open* 11: e048144

Marsden T, McCartan N, Hadley J, Tuck S, Brown L, Haire AJ, Moss CL, Green S, Van Hemelrijck M, Coolen T, *et al* (2021b) Update from the ReIMAGINE Prostate Cancer Screening Study NCT04063566: Inviting Men for Prostate Cancer Screening Using Magnetic Resonance Imaging. *Eur Urol Focus* doi:10.1016/j.euf.2021.03.027 [PREPRINT]

Mazzone E, Stabile A, Pellegrino F, Basile G, Cignoli D, Cirulli GO, Sorce G, Barletta F, Scuderi S, Bravi CA, *et al* (2021) Positive Predictive Value of Prostate Imaging Reporting and Data System Version 2 for the Detection of Clinically Significant Prostate Cancer: A Systematic Review and Meta-analysis. *Eur Urol Oncol* 4: 697–713

Meyer HJ, Wienke A & Surov A (2020) Discrimination between clinical significant and insignificant prostate cancer with apparent diffusion coefficient - A systematic review and meta analysis. *BMC Cancer* 20

Mottet N, Bellmunt J, Briers E, Bergh RCN van den, Bolla M, Casteren NJ van, Conford P, Culine S, Joniau S, Lam T, *et al* (2020) EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer. *Update*

Mottet N, van den Bergh RCN, Briers E, Cornford P, De Santis M, Bellmunt J, Bolla M, van Casteren NJ, Cornford P, Culine S, *et al* (2019) Guidelines on Prostate Cancer 2019. *Eur Assoc Urol Guidel*

Muller BG, Shih JH, Sankineni S, Marko J, Rais-Bahrami S, George AK, de la Rosette JJMCH, Merino MJ, Wood BJ, Pinto P, *et al* (2015) Prostate Cancer: Interobserver Agreement and Accuracy with the Revised Prostate Imaging Reporting and Data System at Multiparametric MR Imaging. *Radiology*

Nam RK, Wallis CJD, Stojcic-Bendavid J, Milot L, Sherman C, Sugar L & Haider MA

(2016) A Pilot Study to Evaluate the Role of Magnetic Resonance Imaging for Prostate Cancer Screening in the General Population. *J Urol* 196: 361–366

NICE Guidance - Prostate cancer: diagnosis and management: © NICE (2019) Prostate cancer: diagnosis and management (2019) *BJU Int*

Norris JM, Carmona Echeverria LM, Bott SRJ, Brown LC, Burns-Cox N, Dudderidge T, El-Shater Bosaily A, Frangou E, Freeman A, Ghei M, *et al* (2020) What Type of Prostate Cancer Is Systematically Overlooked by Multiparametric Magnetic Resonance Imaging? An Analysis from the PROMIS Cohort. *Eur Urol* 78: 163–170

Obmann VC, Pahwa S, Tabayayong W, Jiang Y, O'Connor G, Dastmalchian S, Lu J, Shah S, Herrmann KA, Paspulati R, *et al* (2018) Diagnostic Accuracy of a Rapid Biparametric MRI Protocol for Detection of Histologically Proven Prostate Cancer. *Urology* 122: 133–138

Park KJ, Choi SH, Lee JS, Kim JK & Kim MH (2020) Interreader Agreement with Prostate Imaging Reporting and Data System Version 2 for Prostate Cancer Detection: A Systematic Review and Meta-Analysis. *J Urol* 204: 661–670

Penzkofer T, Padhani AR, Turkbey B, Haider MA, Huisman H, Walz J, Salomon G, Schoots IG, Richenberg J, Villeirs G, *et al* (2021) ESUR/ESUI position paper: developing artificial intelligence for precision diagnosis of prostate cancer using magnetic resonance imaging. *Eur Radiol* 31: 9567–9578

Pickersgill NA, Vetter JM, Andriole GL, Shetty AS, Fowler KJ, Mintz AJ, Siegel CL & Kim EH (2018) Accuracy and Variability of Prostate Multiparametric Magnetic Resonance Imaging Interpretation Using the Prostate Imaging Reporting and Data System: A Blinded Comparison of Radiologists. *Eur Urol Focus*

Rais-Bahrami S, Siddiqui MM, Vourganti S, Turkbey B, Rastinehad AR, Stamatakis L, Truong H, Walton-Diaz A, Hoang AN, Nix JW, *et al* (2015) Diagnostic value of biparametric magnetic resonance imaging (MRI) as an adjunct to prostate-specific antigen (PSA)-based detection of prostate cancer in men without prior biopsies. *BJU Int* 115: 381–388

Roobol MJ, Verbeek JFM, van der Kwast T, Kümmerlin IP, Kweldam CF & van Leenders GJLH (2017) Improving the Rotterdam European Randomized Study of Screening for Prostate Cancer Risk Calculator for Initial Prostate Biopsy by Incorporating the 2014 International Society of Urological Pathology Gleason Grading and Cribriform growth. *Eur Urol*

Rosario DJ, Lane JA, Metcalfe C, Donovan JL, Doble A, Goodwin L, Davis M, Catto JWF, Avery K, Neal DE, *et al* (2012) Short term outcomes of prostate biopsy in men tested for cancer by prostate specific antigen: Prospective evaluation

within ProtecT study. *BMJ* 344

Rosenkrantz AB, Ginocchio LA, Cornfeld D, Froemming AT, Gupta RT, Turkbey B, Westphalen AC, Babb JS & Margolis DJ (2016) Interobserver Reproducibility of the PI-RADS Version 2 Lexicon: A Multicenter Study of Six Experienced Prostate Radiologists. *Radiology*

Rouvière O, Puech P, Renard-Penna R, Claudon M, Roy C, Mège-Lechevallier F, Decaussin-Petrucci M, Dubreuil-Chambardel M, Magaud L, Remontet L, *et al* (2019) Use of prostate systematic and targeted biopsy on the basis of multiparametric MRI in biopsy-naive patients (MRI-FIRST): a prospective, multicentre, paired diagnostic study. *Lancet Oncol* 20: 100–109

Sabouri S, Chang SD, Savdie R, Zhang J, Jones EC, Goldenberg SL, Black PC & Kozlowski P (2017) Luminal water imaging: A new MR imaging T2 mapping technique for prostate cancer diagnosis. *Radiology* 284: 451–459

Schimmöller L, Quentin M, Arsov C, Lanzman RS, Hiester A, Rabenalt R, Antoch G, Albers P & Blondin D (2013) Inter-reader agreement of the ESUR score for prostate MRI using in-bore MRI-guided biopsies as the reference standard. *Eur Radiol*

Schoots IG, Barentsz JO, Bittencourt LK, Haider MA, Macura KJ, Margolis DJA, Moore CM, Oto A, Panebianco V, Siddiqui MM, *et al* (2021) PI-RADS Committee Position on MRI Without Contrast Medium in Biopsy-Naive Men With Suspected Prostate Cancer: Narrative Review. *AJR Am J Roentgenol* 216: 3–19

Shaish H, Kang SK & Rosenkrantz AB (2017) The utility of quantitative ADC values for differentiating high-risk from low-risk prostate cancer: a systematic review and meta-analysis. *Abdom Radiol* 42: 260–270

Shih JH, Greer MD & Turkbey B (2018) The Problems with the Kappa Statistic as a Metric of Interobserver Agreement on Lesion Detection Using a Third-reader Approach When Locations Are Not Prespecified. *Acad Radiol*

Simmons LAM, Ahmed HU, Moore CM, Punwani S, Freeman A, Hu Y, Barratt D, Charman SC, der Meulen J & Emberton M (2014) The PICTURE study -- prostate imaging (multi-parametric MRI and Prostate HistoScanning™) compared to transperineal ultrasound guided biopsy for significant prostate cancer risk evaluation. *Contemp Clin Trials* 37: 69—83

Simmons LAM, Kanthabalan A, Arya M, Briggs T, Barratt D, Charman SC, Freeman A, Gelister J, Hawkes D, Hu Y, *et al* (2017) The PICTURE study: Diagnostic accuracy of multiparametric MRI in men requiring a repeat prostate biopsy. *Br J Cancer* 116: 1159–1165

Smith CP, Harmon SA, Barrett T, Bittencourt LK, Law YM, Shebel H, An JY,

Czarniecki M, Mehralivand S, Coskun M, *et al* (2019) Intra- and interreader reproducibility of PI-RADSv2: A multireader study. *J Magn Reson Imaging*

Sonn GA, Fan RE, Ghanouni P, Wang NN, Brooks JD, Loening AM, Daniel BL, To'o KJ, Thong AE & Leppert JT (2018) Prostate Magnetic Resonance Imaging Interpretation Varies Substantially Across Radiologists. *Eur Urol Focus*

Stabile A, Dell'Oglio P, De Cobelli F, Esposito A, Gandaglia G, Fossati N, Brembilla G, Cristel G, Cardone G, Deho' F, *et al* (2018) Association Between Prostate Imaging Reporting and Data System (PI-RADS) Score for the Index Lesion and Multifocal, Clinically Significant Prostate Cancer. *Eur Urol Oncol* 1: 29–36

The National Lung Screening Trial Research Team (2013) Results of Initial Low-Dose Computed Tomographic Screening for Lung Cancer. *N Engl J Med* 368: 1980–1991

Thompson IM, Ankerst DP, Chi C, Lucia MS, Goodman PJ, Crowley JJ, Parnes HL & Coltman CA (2005) Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/mL or lower. *J Am Med Assoc* 294: 66–70

Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, Minasian LM, Ford LG, Lippman SM, Crawford ED, *et al* (2004) Prevalence of Prostate Cancer among Men with a Prostate-Specific Antigen Level ≤4.0 ng per Milliliter. *N Engl J Med* 350: 2239–2246

Thurfjell EL, Lernevall KA & Taube AAS (1994) Benefit of independent double reading in a population-based mammography screening program. *Radiology* 191: 241–244

Turkbey B, Rosenkrantz AB, Haider MA, Padhani AR, Villeirs G, Macura KJ, Tempany CM, Choyke PL, Cornud F, Margolis DJ, *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* doi:10.1016/j.eururo.2019.02.033 [PREPRINT]

Vis AN, Roemeling S, Kranse R, Schröder FH & van der Kwast TH (2007) Should We Replace the Gleason Score with the Amount of High-Grade Prostate Cancer? *Eur Urol* 51: 931–939

Wang W, Davis CS & Soong SJ (2006) Comparison of predictive values of two diagnostic tests from the same sample of subjects using weighted least squares. *Stat Med* 25: 2215–2229 doi:10.1002/sim.2332 [PREPRINT]

Weinreb JC, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, Margolis D, Schnall MD, Shtern F, Tempany CM, *et al* (2016) PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. *Eur Urol*

Weiss J, Martirosian P, Notohamiprodjo M, Kaufmann S, Othman AE, Grosse U, Nikolaou K & Gatidis S (2018) Implementation of a 5-Minute Magnetic Resonance Imaging Screening Protocol for Prostate Cancer in Men with Elevated Prostate-Specific Antigen before Biopsy. *Invest Radiol* 53: 186–190

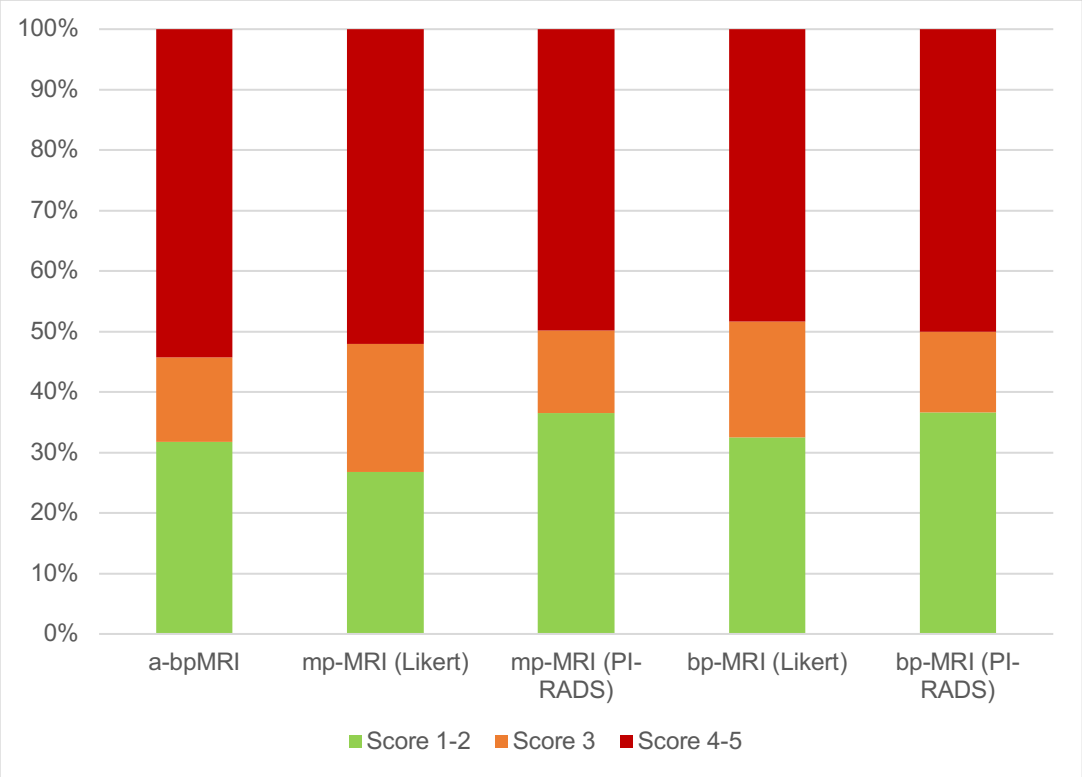
Westphalen AC, McCulloch CE, Anaokar JM, Arora S, Barashi NS, Barentsz JO, Bathala TK, Bittencourt LK, Booker MT, Braxton VG, *et al* (2020) Variability of the Positive Predictive Value of PI-RADS for Prostate MRI across 26 Centers: Experience of the Society of Abdominal Radiology Prostate Cancer Disease-focused Panel. *Radiology* 296: 76–84

Wolters T, Roobol MJ, van Leeuwen PJ, van den Bergh RCN, Hoedemaeker RF, van Leenders GJLH, Schröder FH & van der Kwast TH (2010) Should Pathologists Routinely Report Prostate Tumour Volume? The Prognostic Value of Tumour Volume in Prostate Cancer. *Eur Urol* 57: 821–829

Zawaideh JP, Sala E, Shaida N, Koo B, Warren AY, Carmisciano L, Saeb-Parsy K, Gnanapragasam VJ, Kastner C & Barrett T (2020) Diagnostic accuracy of biparametric versus multiparametric prostate MRI: assessment of contrast benefit in clinical practice. *Eur Radiol* 30: 4039–4049

Appendix

Supplementary Figure 1. Proportions of MRI scores.



MRI scores were grouped as follows: score 1-2, 3 and 4-5. MRI score distribution for a-bpMRI was 32%, 14% and 54%, respectively; for mp-MRI Likert scores was 27%, 21% and 52%, respectively; for mp-MRI PI-RADS scores was 36%, 14% and 50%; for bp-MRI Likert scores was 32%, 19% and 48%, respectively; for bp-MRI PI-RADS scores was 37%, 13% and 50%. Proportions are on the total number of patients included in the study (n = 151).

Supplementary Table 1. MRI protocol details

Sequence	TR (ms)	TE (ms)	FoV (mm)	Slice thickness (mm)	Gap (mm)
T2 TSE coronal	6128	100	180	3	3
T2 TSE axial	5407	100	180	3	0
T1W TSE	487	8	240	3	3
DWI 0 150 500 1000	2753	80	220	5	0
DWI b2000	2000	78	220	5	0
DCE	5.8	2.8	180	3	0

Supplementary Table 2. Diagnostic accuracy of MRI.

	a-bpMRI	bp-MRI (Likert)	bp-MRI (PI-RADS)	mp-MRI (Likert)	mp-MRI (PI-RADS)
Gleason ≥3+4 and/or MCCL ≥4 mm (definition 2); prevalence 63%					
Sensitivity	87 (83-91)	89 (85-93)	84 (79-88)	89 (85-93)	84 (79-88)
Specificity	65 (57-72)*	55 (48-63)	70 (62-76)	54 (46-62)	71 (64-78)
PPV	81 (76-85)	77 (72-82)	82 (78-87)	77 (72-81)	83 (79-87)
NPV	75 (67-82)	75 (66-82)	72 (65-79)	75 (67-83)	73 (65-79)
Any Gleason ≥3+4; prevalence 50%					
Sensitivity	88 (83-92)	90 (86-94)	86 (81-90)	91 (86-94)	86 (81-90)
Specificity	52 (45-59) * ‡	44 (39-52)	58 (51-64)	44 (38-51)	59 (52-66)
PPV	65 (59-70)	63 (57-68)	67 (62-73)	62 (57-68)	68 (62-73)
NPV	81 (73-87)	82 (74-89)	80 (73-86)	83 (75-89)	81 (74-86)

Pooled values (from the three readers) are reported as % (95%-CI). *p<0.05 a-bpMRI vs bp/mp-MRI Likert. ‡ p<0.05 a-bpMRI vs bp/mp-MRI PI-RADS. bp-MRI: biparametric MRI; mp-MRI: multiparametric MRI; a-bpMRI: abbreviated biparametric MRI; PPV: positive predictive value; NPV: negative predictive value; MCCL: maximum cancer core length.

Supplementary table 3. Diagnostic accuracy of a-bpMRI (combined scores)

	Definition 2 csPCa*	***Any Gleason ≥3+4***
	MRI score ≥4	
Sens	76 (67-84)	78 (68-87)
Spec	80 (70-91)	68 (57-79)
PPV	87 (79-94)	71 (61-81)
NPV	66 (55-77)	75 (65-85)
	T2 and DWI score ≥4	
Sens	60 (50-70)	62 (51-73)
Spec	91 (84-99)	80 (71-89)
PPV	91 (85-99)	76 (65-86)
NPV	57 (47-68)	67 (58-77)

Values are reported as % (95%-CI). *Gleason ≥3+4 and/or MCCL ≥4 mm. PPV: positive predictive value; NPV: negative predictive value

Supplementary Table 4. NPV and positivity rates of abbreviated bp-MRI (a-bpMRI) according to clinically significant PCa prevalence.

	Prevalence of csPCa		
	10%	**5%**	**2%**
Definition 2 csPCa*			
MRI score ≥3			
Pos. rate	42 (34-50)	38 (31-47)	37 (29-45)
PPV	22 (13-34)	12 (5-23)	5 (1-15)
NPV	99 (94-100)	99 (94-100)	100 (96-100)
MRI score ≥4			
Pos. rate	25 (18-33)	23 (16-30)	21 (14-28)
PPV	29 (15-46)	18 (7-35)	6 (1-21)
NPV	96 (91-99)	98 (94-100)	99 (95-100)
T2WI and DWI score ≥4			
Pos. rate	14 (9-20)	12 (7-18)	10 (6-16)
PPV	43 (22-66)	28 (10-53)	13 (2-40)
NPV	95 (90-98)	98 (94-100)	99 (96-100)
Any Gleason ≥3+4			
MRI score ≥3			
Pos. rate	52 (44-60)	50 (42-59)	49 (51-57)
PPV	18 (10-28)	9 (4-18)	4 (1-11)
NPV	99 (93-100)	100 (95-100)	100 (95-100)
MRI score ≥4			
Pos. rate	37 (29-45)	34 (27-42)	32 (25-41)
PPV	22 (12-35)	12 (4-23)	4 (0-14)
NPV	97 (91-99)	98 (93-100)	99 (95-100)
T2WI and DWI score ≥4			
Pos. rate	24 (17-31)	22 (16-30)	21 (15-29)
PPV	25 (12-42)	15 (5-31)	6 (1-21)
NPV	95 (89-98)	97 (93-99)	99 (95-100)

Values are reported as % (95%-CI). *Gleason ≥3+4 and/or MCCL ≥4 mm. Pos. rate: rate of positive test according to MRI cut-off. csPCa: clinically significant prostate cancer; pos. rate: positivity rate of MRI; PPV: positive predictive value; NPV: negative predictive value.

Supplementary Table 5. Clinically significant lesions (definition 1) missed by MRI.

	Mp-MRI (Likert)	Mp-MRI (PI-RADS)	a-bpMRI
	Number (MCCL range, mm)		
MRI score ≥3			
Gleason 3+3	2 (7-10)	4 (6-10)	2 (6-10)
Gleason 3+4	1 (6)	2 (6-10)	0
Gleason 4+3	1 (4)	1 (4)	1 (4)
MRI score ≥4			
Gleason 3+3	4 (6-10)	6 (6-10)	5 (6-10)
Gleason 3+4	5 (6-10)	6 (6-10)	3 (6-10)
Gleason 4+3	1 (4)	2 (4-8)	1 (4)
T2WI and DWI score ≥4			
Gleason 3+3	na	na	7 (6-10)
Gleason 3+4	na	na	7 (6-12)
Gleason 4+3	na	na	3 (4-8)

The table refers to MRI lesions missed by at least 2 out of 3 readers using different cut-off for MRI positivity. MCCL: maximum cancer core length; na: not assessed.

Supplementary Table 6. Interreader agreement (cut off: MRI score ≥3).

	Mp-MRI (Likert)	*Mp-MRI (PI-RADS)*	*Bp-MRI (Likert)*	*Bp-MRI (PI-RADS)*	*a-bpMRI*
Gwet's AC1	0.59 (0.48-0.69)	0.65 (0.56-0.75)	0.58 (0.48-0.69)	0.61 (0.51-0.71)	0.58 (0.48-0.69)
Percentage of agreement	75% (70-80)	81% (77-86)	75% (70-80)	79% (74-84)	76% (72-82)

Gwet's AC1: agreement coefficient which takes values between 0 to 1, similar to kappa. Values in parentheses are 95%-CI.
mp-MRI: multiparametric MRI; a-bpMRI: abbreviated biparametric MRI.

Supplementary Table 7. Interreader agreement of a-bpMRI using alternative MRI cut-offs.

	MRI score ≥4	T2 ≥4 and DWI≥4
Gwet's AC1	0.59 (0.49-0.69)	0.64 (0.53-0.72)
Percentage of agreement	79% (74-84)	82% (77-87)

Gwet's AC1: agreement coefficient which takes values between 0 to 1, similar to kappa. Values in parentheses are 95%-CI.