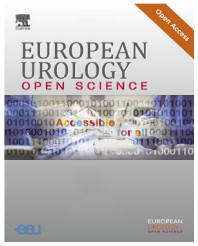
available at www.sciencedirect.com
journal homepage: www.eu-openscience.europeanurology.com



Prostate Cancer

Inter-reader Agreement for Prostate Cancer Detection Using Micro-ultrasound: A Multi-institutional Study

Steve R. Zhou [a,*], *Moon Hyung Choi* [a,b,c], *Sulaiman Vesal* [a,b], *Adam Kinnaird* [d], *Wayne G. Brisbane* [e], *Giovanni Lughezzani* [f,g], *Davide Maffei* [f,g], *Vittorio Fasulo* [f,g], *Patrick Albers* [d], *Lichun Zhang* [b], *Zachary Kornberg* [a], *Richard E. Fan* [a], *Wei Shao* [h], *Mirabela Rusu* [a,b,i,†], *Geoffrey A. Sonn* [a,b,†]

[a] *Department of Urology, Stanford School of Medicine, Palo Alto, CA, USA;* [b] *Department of Radiology, Stanford School of Medicine, Palo Alto, CA, USA;* [c] *Department of Radiology, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea;* [d] *Department of Urology, University of Alberta, Edmonton, Canada;* [e] *Department of Urology, University of California-Los Angeles, Los Angeles, CA, USA;* [f] *Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy;* [g] *Department of Urology, IRCCS Humanitas Research Hospital, Milan, Italy;* [h] *Department of Medicine, University of Florida, Gainesville, FL, USA;* [i] *Department of Biomedical Data Science, Stanford School of Medicine Palo Alto, CA, USA*

Article info

Article history:
Accepted June 19, 2024

Associate Editor:
Roderick van den Bergh

Keywords:
Inter-reader agreement
Micro-ultrasound
Prostate cancer
Targeted biopsy

Abstract

Background and objective: Micro-ultrasound (MUS) uses a high-frequency transducer with superior resolution to conventional ultrasound, which may differentiate prostate cancer from normal tissue and thereby allow targeted biopsy. Preliminary evidence has shown comparable sensitivity to magnetic resonance imaging (MRI), but consistency between users has yet to be described. Our objective was to assess agreement of MUS interpretation across multiple readers.

Methods: After institutional review board approval, we prospectively collected MUS images for 57 patients referred for prostate biopsy after multiparametric MRI from 2022 to 2023. MUS images were interpreted by six urologists at four institutions with varying experience (range 2–6 yr). Readers were blinded to MRI results and clinical data. The primary outcome was reader agreement on the locations of suspicious lesions, measured in terms of Light's κ and positive percent agreement (PPA). Reader sensitivity for identification of grade group (GG) ≥2 prostate cancer was a secondary outcome.

Key findings and limitations: Analysis revealed a κ value of 0.30 (95% confidence interval [CI] 0.21–0.39). PPA was 33% (95% CI 25–42%). The mean patient-level sensitivity for GG ≥2 cancer was 0.66 ± 0.05 overall and 0.87 ± 0.09 when cases with anterior lesions were excluded. Readers were 12 times more likely to detect higher-grade cancers (GG ≥3), with higher levels of agreement for this subgroup (κ 0.41, PPA 45%). Key limitations include the inability to prospectively biopsy reader-delineated targets and the inability of readers to perform live transducer maneuvers.

[†] These authors contributed equally as senior authors.

[*] Corresponding author. Department of Urology, Stanford School of Medicine, 453 Quarry Road, Palo Alto, CA 94304, USA. Tel. +1 650 725 5746; Fax: +1 650 498 5346.
E-mail address: steveranzhou@gmail.com (S.R. Zhou).

https://doi.org/10.1016/j.euros.2024.06.017

Conclusions and clinical implications: Inter-reader agreement on the location of suspicious lesions on MUS is lower than rates previously reported for MRI. MUS sensitivity for cancer in the anterior gland is lacking.

Patient summary: The ability to find cancer on imaging scans can vary between doctors. We found that there was frequent disagreement on the location of prostate cancer when doctors were using a new high-resolution scan method called micro-ultrasound. This suggests that the performance of micro-ultrasound is not yet consistent enough to replace MRI (magnetic resonance imaging) for diagnosis of prostate cancer.

1. Introduction

Despite many recent advances, prostate cancer (PCa) diagnosis can still be improved. While randomized trials have demonstrated that many patients with PCa can be managed safely with active surveillance, identification of these men hinges on accurate risk stratification via imaging-guided biopsy. This task has been limited by the current state of medical imaging [1,2]. Because conventional transrectal ultrasound (TRUS) cannot reliably differentiate cancer from normal prostatic tissue, biopsies traditionally involved systematic sampling of the gland, an approach that underestimates cancer risk while overdetecting low-grade cancers that do not benefit from treatment [3,4]. The development of multiparametric magnetic resonance imaging (mpMRI) has allowed clinicians to perform targeted biopsy of suspicious lesions. While this has improved the detection of clinically significant PCa [3,4], it is expensive, has medical contraindications, needs radiology expertise that is not widespread, and requires accurate co-registration between MRI and ultrasound [5–7].

Micro-ultrasound (MUS) uses a high-frequency (29 MHz) transducer with superior resolution to conventional TRUS (6–10 MHz) [8]. MUS may differentiate PCa from normal tissue, which could allow live targeted biopsy. Preliminary evidence has shown that MUS has comparable sensitivity to mpMRI at high-volume centers, but consistency between users has yet to be described [9–11]. Variability in image interpretation is a known limitation of mpMRI [12–14]. We therefore sought to characterize inter-reader agreement of MUS interpretation among readers at different institutions.

2. Patients and methods

2.1. Study population

After obtaining institutional review board approval, we prospectively enrolled patients undergoing mpMRI and prostate biopsy at a single institution for elevated prostate-specific antigen (PSA) or continued active surveillance from 2022 to 2023. To match cancer prevalence to a typical biopsy population, we included additional patients with negative biopsy from a second institution (University of Florida) [3,4]. Informed consent was obtained from all subjects.

2.2. Imaging and biopsy procedure

All patients underwent mpMRI performed with a 3.0-T scanner without an endorectal coil in accordance with a Prostate Imaging-Reporting and Data System (PI-RADS)-compliant protocol [15]. MRI scans were interpreted and all suspicious lesions (PI-RADS version 2.1 score $\geq$3) were outlined by fellowship-trained radiologists. Two urologists (Stanford and University of Florida) performed targeted biopsy of MRI-visible lesions under cognitive fusion, combined with 12–14-core systematic biopsy. At least two targeted cores were obtained for each target. All systematic locations were sampled even if they overlapped with the mpMRI target. Biopsies were performed via a transperineal approach using a Precision Point platform (Perineologic, Cumberland, MD, USA) and a BK Ultrasound unit (BK Medical, Burlington, MA, USA).

Immediately before biopsy, MUS images of the prostate were obtained in a single sweep from right to left using the ExactVu transrectal MUS system (Exact Imaging Inc., Markham, ON, Canada). MUS was not used to guide additional biopsies. Images were reconstructed in three-dimensional (3D) format and saved for later review. We excluded images with insufficient quality or field of view.

2.3. Study design

Six urologists at four institutions with varying degrees of MUS experience (range 2–6 yr) were provided with image stacks for all patients and asked to annotate any suspicious lesions (PRI-MUS score $\geq$3) using 3D Slicer [16]. All readers underwent comprehensive online training on the PRI-MUS protocol [17]. Readers could scroll through the entire image stack for a given patient in a similar fashion to a live ultrasound sweep in clinic. Readers were not instructed to assign a specific PRI-MUS score because score agreement was not analyzed. They were blinded to PSA, age, MRI results, prior biopsy history, and pathology results.

The primary outcome was inter-reader agreement on the locations of suspicious lesions on MUS. Reader sensitivity for detection of biopsy-proven grade group (GG) $\geq$2 PCa on MUS was a secondary outcome.

2.4. Generation of reference lesions

MUS images, biopsy pathology, and mpMRI results were cross-referenced by a urologist and genitourinary radiolo-

gist together to annotate areas on MUS corresponding to biopsy-proven GG ≥2 PCa. These areas were termed "reference lesions" and fell into two categories: MRI-visible and MRI-invisible PCa.

For MRI-visible lesions, we identified the reference lesion by performing software co-registration to transfer radiologist annotations from the MRI onto the MUS image stack (Fig. 1). For MRI-invisible PCa, the MUS images were cross-matched cognitively with the template locations of positive systematic cores to localize the reference lesion within the corresponding anatomic sector.

2.5. Statistical analysis

Inter-reader agreement was measured using Light's κ, which is a variation of Cohen's κ appropriate for fully crossed studies with multiple readers [18–21]. Agreement analysis did not depend on biopsy pathology results. We used standard measures to mitigate κ underestimation in the presence of high marginal frequencies as outlined by Shih et al [22]. First, we reported positive percent agreement (PPA) [23]. Second, κ was calculated using sector-based analysis; each prostate was segmented on 3D Slicer and computationally partitioned into 30 sectors on the basis of transperineal biopsy templates (Fig. 2A). Reader annotations were considered concordant if they occupied the same sector(s) of a given gland; a sector was considered occupied if >20% of an annotation volume fell within that sector, and—to account for large lesions—if >50% of the sector volume was occupied by an annotation. The 20% cutoff was determined on the basis of the maximum number of sectors that can describe the anatomic location of a lesion (n_sectors = 4). The threshold would be $1/(n\_\text{sectors} + 1) = 0.20$. Multiple agreements were allowed for annotations occupying more than one sector (Fig. 2B). Confidence intervals (CIs) for κ were generated as described by Fleiss et al [24]. CIs for PPA were generated using the Delta method [25].

For analysis of secondary outcomes, reader annotations were considered true positives if there was >20% overlap with a reference lesion. Summary statistics were computed at the lesion and patient levels. Owing to over-representation of true-negative prostate tissue by volume, the lesion-level specificity and negative predictive value (NPV) are not clinically interpretable and are not reported.

Image analysis was performed using the *SimpleITK* Python package (Python Software Foundation, Wilmington, DE, USA). Statistical analysis was performed using the *statsmodel.api* Python package (Python Software Foundation). Figures were created with BioRender.com.

2.6. Subgroup analyses

We investigated whether inter-reader agreement was higher in specific patient subgroups. Clinical risk factors assessed include positive versus negative biopsy, higher cancer grade (GG ≥3), PSA ≥10 ng/ml or PSA density (PSAD) ≥0.15 ng/ml/cm^3, and biopsy history (active surveillance, prior negative biopsy, or biopsy-naïve). Anatomic risk factors included large prostate size (volume ≥40 cm^3), larger lesions, and lesion location (anterior included vs posterior only). Owing to the angular sagittal acquisition space for MUS, voxel dimensions vary from case to case and do not represent true lesion volume in 3D space, instead serving as comparative approximations. We split cases into two subgroups using the median lesion volume measured in voxels. Owing to the presence of MRI-negative lesions, MRI lesion volume could not be used as a surrogate. We also assessed whether readers with ≥4 yr of experience in interpreting MUS had higher agreement for cancer detection than readers with less experience. We conducted multivariable and univariable logistic regression to evaluate whether these factors would increase the odds that readers would successfully identify GG ≥2 PCa.

3. Results

3.1. Patient population

Comprehensive demographics are listed in Table 1. The data set included 57 patients with a median age of 68 yr (interquartile range [IQR] 61–71) and median PSA of 6.2 ng/ml (IQR 4.5–8.1). In total, 26 patients (47%) had biopsy-proven GG ≥2 PCa. There were 30 MUS reference lesions; 22 patients (39%) had one reference lesion, whereas four (7.0%) harbored two lesions. Twenty-eight reference lesions (93%) were identified from MRI targets, whereas only two (7%) were detected in nonadjacent systematic biopsy cores. Both MRI-negative reference lesions were in the transitional zone and visible on MUS. Only one of these lesions had higher-grade cancer in the nonadjacent systematic core than in the MRI target. Of the MUS reference lesions, nine were located anteriorly (30%) and 21 were posterior (70%).

3.2. Primary outcome

Reader agreement on the location of MUS reference lesions was fair. Given any reader observation in a prostate sector, there was a 30% chance that a second randomly selected reader would make the same observation (κ = 0.30, 95% CI 0.21–0.39). Given any positive reader annotation, there was a 33% chance that a second randomly selected reader would agree on the presence of cancer in that sector (PPA 33%, 95% CI 25–0.42%). Both κ and PPA were consistent across reader pairs (standard deviation [SD] 0.05). Reader 2 annotated the highest number of lesions (n = 70), designating only four (7%) patients as having no cancer. Reader 3 annotated the fewest lesions (n = 37), designating 26 prostates (46%) as normal.

3.3. Secondary outcome

For detection of GG ≥2 PCa at a patient level, the mean sensitivity was 0.66 (SD 0.05) and the mean specificity was 0.37 (SD 0.17) across the six readers (Table 2). The mean positive predictive value (PPV) was 0.47 (SD 0.05) and the mean NPV was 0.53 (SD 0.14). At a lesion level, the mean sensitivity was 0.58 (SD 0.05) and the mean PPV was 0.34 (SD 0.06). In total, seven lesions (23%) were missed by all six readers, whereas ten lesions (33%) were detected by all six readers. Combining reader annotations resulted in aggregate lesion-level sensitivity of 0.77 (23/30 lesions) and patient-level sensitivity of 0.85 (22/26 patients).

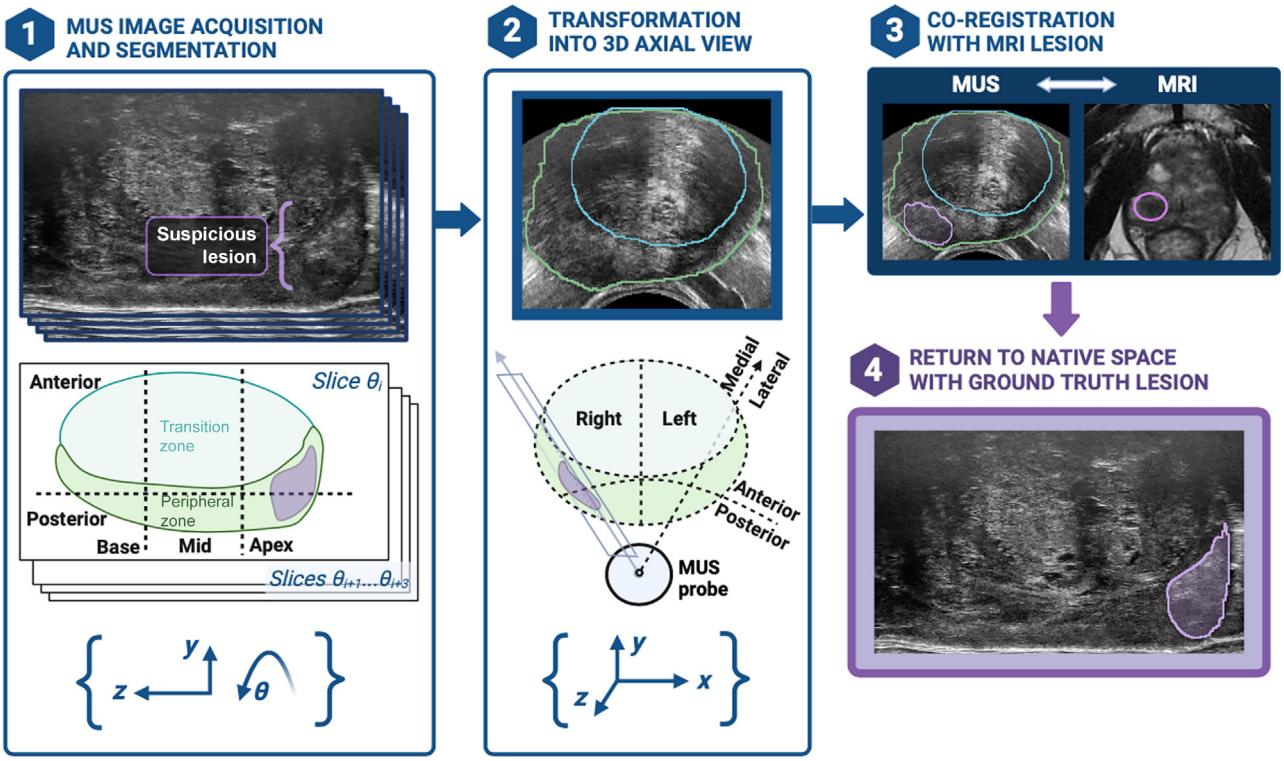


Fig. 1 – Image analysis pipeline for mapping the location of grade group ≥2 prostate cancer on micro-ultrasound (MUS) images using magnetic resonance imaging (MRI) and biopsy information. (1) MUS image stacks are captured in the native angular sagittal space (*θ, y, z*) and the gland components are segmented. (2) Native-space images are transformed into a three-dimensional (3D) Cartesian space (*x, y, z*) to allow co-registration of the axial MUS with axial MRI. (3) The radiologist's annotations of biopsy-confirmed MRI-visible lesions are co-registered onto the axial-view MUS image stack as 3D segmentations. (4) Lesion segmentations are "de-transformed" back into the original coordinate system on the MUS image stack (*θ, y, z*). Each de-transformed segmentation is then manually refined to yield a final reference lesion annotation by reviewing and adjusting it to match the corresponding lesion seen on MUS and to encompass adjacent positive cores in the native angular sagittal space. All co-registrations and final reference lesions were manually reviewed by the expert radiologist (M.H.C.) and urologist (S.R.Z.) in conjunction to ensure that the multiparametric MRI lesion and biopsy results corresponded anatomically to the reference lesion seen on MUS and vice versa.

3.4. *Subgroup analyses*

The level of agreement among readers was higher for patients with positive biopsy, higher-grade cancer, higher PSA and PSAD, and larger glands and lesions. Reader agreement tended to be better for patients without anterior lesions (κ = 0.39, 95% CI 0.24–0.53; PPA = 43%, 95% CI 30.7–55.3%), although the wide CIs for patients with anterior lesions suggest sample size limitations (κ = 0.26, 95% CI 0.02–0.50; PPA = 29%, 95% CI 3.2–54.7%). Agreement was not higher among readers with ≥4 yr of experience in interpreting MUS than among readers with less experience. Findings were consistent whether measured using κ or PPA.

Univariable analysis revealed that readers were more than 12 times more likely to detect true positives in patients with higher-grade cancer (GG ≥3). The mean sensitivity for GG ≥3 cases was 0.75 (SD 0.07) at the lesion level and 0.88 (SD 0.10) at the patient level. By contrast, readers were 20 times more likely to miss cancer in cases with anterior lesions. Of the seven lesions missed by all readers, six were anterior. While four of these cases did have corpora amylaceae or periurethral calcifications, none obscured the reference lesion. Exclusion of anterior lesions increased the mean sensitivity to 0.80 (SD 0.08) at the lesion level and 0.87 (SD 0.09) at the patient level. The median prostate volume for cases with anterior lesions missed by the majority of readers was 44.5 cm^3 (IQR 34.4–49.1), which is not notably different from the median prostate volume for the entire cohort (44.3 cm^3, IQR 36.0–58.0). Effects persisted on multivariable analysis. By contrast, PSA, PSAD, and larger lesion size were associated with detection on univariable analysis only. Neither more experience nor larger prostate size increased the likelihood of cancer detection (Table 3).

4. Discussion

Our study produced three key findings and suggests that more work is needed before the MUS performance can approach that of MRI.

First, MUS interpretation consistency across readers appears to be lower than for prostate MRI. In our study, six experienced readers agreed on lesion location on MUS just 33% of the time (PPA = 33%, κ = 0.30). By contrast, Brembilla et al [12] reported PPA of 78.4% and κ of 0.591 among 11 radiologists identifying the index lesion on 132 MRI scans. Accounting for all lesions within the prostate leads to lower MRI agreement, but this is still higher than agreement on MUS. A study by Kohestani et al [14] found κ of 0.41 for three radiologists annotating all lesions on 97 MRI scans using a 24-sector template for a prostatectomy population. In an analysis of nine radiologists inter-

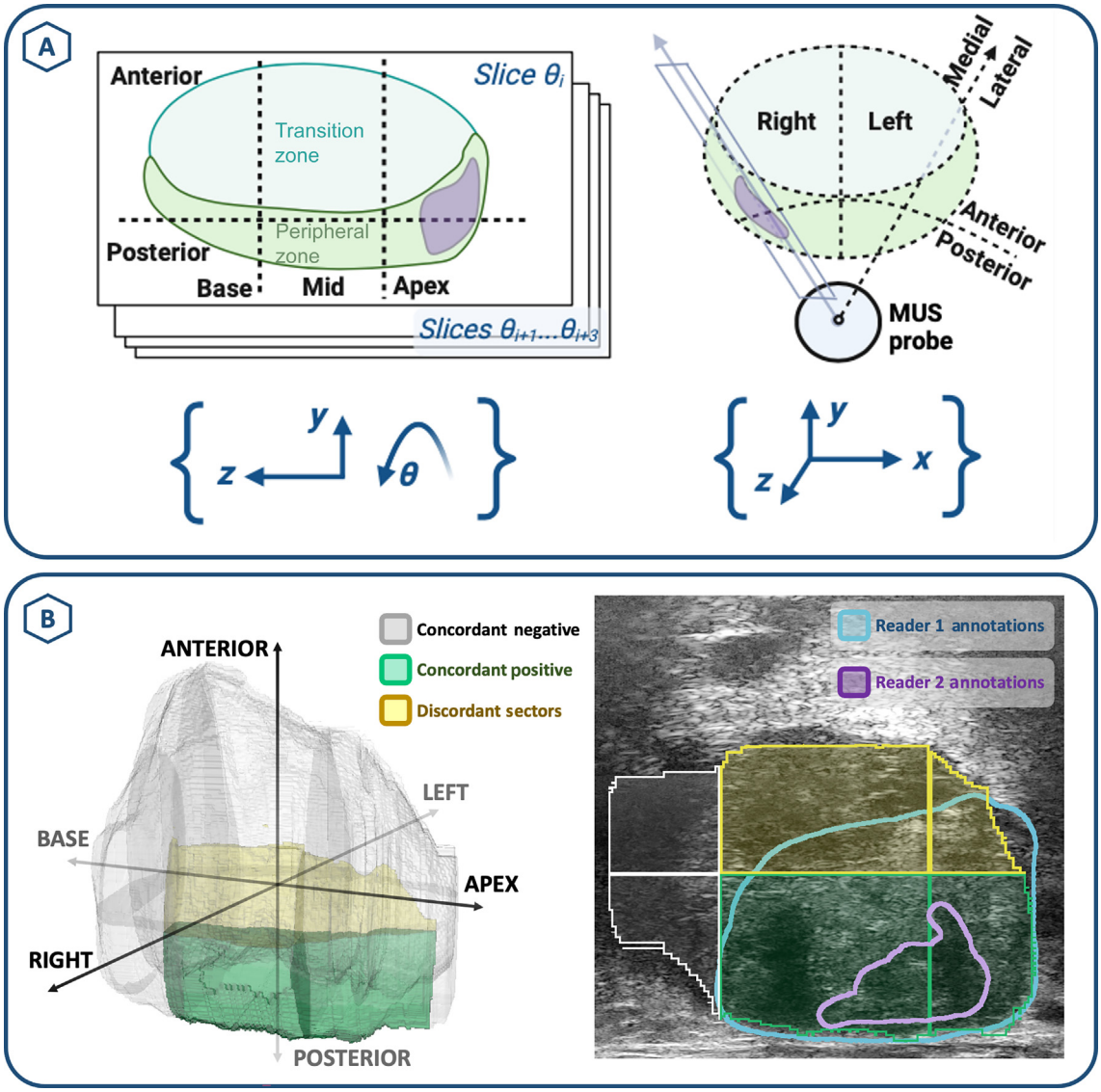


Fig. 2 – Diagram of sector-based agreement analysis. (A) Diagram of how prostate sectors were partitioned. The transition zone was divided into left, right, base, mid, and apex sectors. The peripheral zone was additionally divided into medial, lateral, anterior, and posterior sectors for a total of 30 sectors. (B) Example of a case with partial agreement between two readers. While annotations by both readers occupy the right posterolateral mid and apical sectors of the peripheral zone (green, concordant positive), one annotation also extends into the right anterolateral mid to apex sector (yellow, discordant). Both readers agree on the absence of lesions in the remaining 26 sectors (gray, concordant negative). The biopsy result was negative.

preting 110 MRI scans using a 32-sector template for a prostatectomy population, Greer et al [13] reported PPA of 87.1% for localizing the index lesion, and 56.9% for localizing all lesions. In contrast to results for MRI in the literature, inter-reader agreement in our study was not higher among users with more experience.

Second, our MUS sensitivity of 0.66 is much lower than results from previous series, which range from 0.94 to 1.0 [26]. This difference is largely driven by lesion location; 30% of the lesions in our cohort were in the anterior prostate. Of the nine anterior lesions in our study, six were missed by all readers. By contrast, a 104-patient cohort described by G.L. only included two anterior cases (6%), which were the only two lesions missed. This is consistent with findings that up to 62% of lesions missed on MUS are in the anterior zone, probably owing to the heterogeneous appearance of the transition zone, sources of artifacts such as periurethral calcifications and corpora amylaceae, and overall signal attenuation at greater tissue depths [9]. The PRI-MUS protocol was only recently updated to include criteria for localizing anterior lesions, and the update has yet to undergo prospective external validation [27].

This introduces our third key finding: the lowest agreement and sensitivity were observed for cases with anterior lesions and lower-grade cancer. Exclusion of anterior cases increased the mean reader sensitivity from 0.66 to 0.87 and increased the PPA from 29% to 43%. However, the relative proportion of anterior PCa cases is approximately 30%, so good performance in this anatomic zone remains vital [28]. Subgroup analyses also revealed greater reader agreement for cases with higher-grade cancer. Cases with GG $\geq$3 cancer on biopsy were more than 12 times more likely to be

Table 1 – Summary of cohort demographics

Parameter	Result
Median age, yr (IQR)	68.0 (61.0–71.0)
Median PSA, ng/ml (IQR)	6.2 (4.5–8.1)
Median prostate volume, cm^3 (IQR)	44.3 (36.0–58.0)
Median frames per patient n (IQR)	143.0 (122.0–164.0)
Highest grade group, *n*/*N* (%)	
0	24/57 (42.1)
1	7/57 (12.3)
2	11/57 (19.3)
3	9/57 (15.8)
4	2/57 (3.5)
5	4/57 (7.0)
PI-RADS score, *n*/*N* (%)	
<3	16/57 (28.1)
3	5/57 (8.8)
4	21/57 (36.8)
5	15/57 (26.3)
Biopsy history, *n*/*N* (%)	
Naïve	42/57 (73.7)
Prior negative	3/57 (5.3)
Surveillance	12/57 (21.1)
Reference lesions per patient, *n*/*N* (%)[a]	
0	31/57 (54.4)
1	22/57 (38.6)
2	4/57 (7.0)
>2	0/57 (0.0)
Reference lesion location, *n*/*N* (%)	
Anterior	9/30 (30.0)
Posterior	21/30 (70.0)

IQR = interquartile range; PI-RADS = Prostate Imaging-Reporting and Data System; PSA = prostate-specific antigen.
[a] Number of distinct grade group ≥2 lesions per patient on micro-ultrasound.

detected by readers than GG 2, with mean reader sensitivity of 0.88.

Our study has five noteworthy limitations. First, we did not prospectively biopsy reader-outlined targets owing to the multireader study design. Additional biopsy of these targets may have increased detection because of the multifocality of PCa [9]. However, all patients in our study underwent thorough 14-core systematic transperineal biopsy that included additional targeted cores from MRI-visible lesions, and thus it is likely that the presence of undetected GG ≥2 lesions was low. Second, readers were blinded to clinical factors such as biopsy history and PSA, which would normally improve assessment of the pretest probability. However, our subgroup analysis demonstrated that neither of these factors was associated with agreement or sensitivity when controlling for pathology results. Third, readers could not perform live MUS transducer maneuvers such as translation or compression to help in navigating shadowing, but they did have the advantage of reviewing images without the time constraints of in-clinic biopsy. Fourth, we did not assess reader agreement on PRI-MUS scores, but we did perform subgroup analyses that showed higher agreement for lesions with higher-grade cancer. Furthermore, PRI-MUS score assignment would not change clinical management, as any lesions scored ≥3 would be biopsied. Finally, the sample size is small. Despite this limitation, owing to the fully crossed study design, the number of cases was sufficient to generate κ and PPA values with reliable CIs.

Notwithstanding these limitations, our study has several unique strengths. This is the first study to characterize inter-reader agreement on MUS interpretation, which is a valuable metric for diagnostic performance because it does not depend on the quality of the ground truth. In the biopsy setting, true cancer burden is underestimated [29]. While prostatectomy specimens may capture the full picture, the advanced cancer burden in this group does not represent the intended biopsy population for the diagnostic tool. Our image analysis methods also improved on approaches used for MRI; whereas prior studies often used screenshots or descriptive language to approximate lesion location, we

Table 2 – Summary of reader annotations and performance

Parameter	Reader						Mean ± SD
	1	2	3	4	5	6	
Experience (yr)	3	3	6	2	6	4	4.0 ± 1.7
Lesions annotated (*n*)	54	70	37	56	60	42	53.2 ± 12.0
Sectors annotated (*n*)	90	110	59	93	114	57	87.2 ± 24.4
Lesions per patient, *n* (%)							
0	9 (16)	4 (7)	26 (46)	16 (28)	16 (28)	21 (37)	15.3 ± 7.9
1	42 (74)	39 (68)	25 (44)	28 (49)	26 (46)	30 (53)	31.7 ± 7.1
2	6 (11)	11 (19)	6 (11)	11 (19)	12 (21)	6 (11)	8.7 ± 2.9
>2	0 (0)	3 (5)	0 (0)	2 (4)	3 (5)	0 (0)	1.3 ± 1.5
Lesion-level analysis							
All lesions							
Sensitivity	0.63	0.60	0.53	0.60	0.63	0.50	0.58 ± 0.05
Positive predictive value	0.35	0.26	0.43	0.32	0.32	0.36	0.34 ± 0.06
Anterior lesions excluded							
Sensitivity	0.79	0.89	0.74	0.84	0.84	0.68	0.80 ± 0.08
Positive predictive value	0.33	0.28	0.42	0.33	0.32	0.37	0.34 ± 0.05
Patient-level analysis							
All cases							
Sensitivity	0.69	0.69	0.62	0.69	0.69	0.58	0.66 ± 0.05
Specificity	0.26	0.10	0.58	0.39	0.39	0.48	0.37 ± 0.17
Positive predictive value	0.44	0.39	0.55	0.49	0.49	0.48	0.47 ± 0.05
Negative predictive value	0.50	0.27	0.64	0.60	0.60	0.58	0.53 ± 0.14
Anterior cases excluded							
Sensitivity	0.82	1.00	0.82	0.94	0.88	0.76	0.87 ± 0.09
Specificity	0.26	0.10	0.58	0.39	0.39	0.48	0.37 ± 0.17
Positive predictive value	0.38	0.38	0.52	0.46	0.44	0.45	0.44 ± 0.05
Negative predictive value	0.73	1.00	0.86	0.92	0.86	0.79	0.86 ± 0.10

SD = standard deviation.

Table 3 – Subgroup comparisons of reader agreement (κ, PPA) and the odds of patient-level cancer detection for risk factors

Parameter	κ	PPA (%)	OR (95% confidence interval)	
			Multivariable	Univariable
Positive biopsy (GG ≥2)	0.36 (0.23–0.48)	39.6 (28.3–50.8)	–	–
Negative biopsy	0.23 (0.09–0.36)	25.7 (12.2–39.3)	–	–
GG ≥3	0.41 (0.26–0.56)	45.2 (31.8–58.7)	4.92 (1.49–16.28)	12.57 (5.61–28.13)
GG <3	0.22 (0.11–0.34)	25.5 (14.4–36.5)	Reference	Reference
Anterior lesion	0.26 (0.02–0.50)	29.0 (3.2–54.7)	0.09 (0.03–0.29)	0.05 (0.02–0.12)
Posterior lesion	0.39 (0.24–0.53)	43.0 (30.7–55.3)	Reference	Reference
PSAD ≥0.15 ng/ml/cm^3	0.34 (0.21–0.46)	37.4 (26.6–48.2)	0.64 (0.17–2.35)	2.28 (1.14–4.54)
PSAD <0.15 ng/ml/cm^3	0.26 (0.12–0.40)	29.0 (14.7–43.3)	Reference	Reference
PSA ≥10 ng/ml	0.40 (0.21–0.59)	45.1 (28.5–61.7)	5.12 (0.92–28.43)	10.16 (2.97–34.79)
PSA <10 ng/ml	0.26 (0.16–0.37)	29.4 (19.3–39.6)	Reference	Reference
Prostate volume ≥40 cm^3	0.33 (0.21–0.44)	35.9 (25.0–46.9)	0.47 (0.14–1.52)	1.20 (0.62–2.35)
Prostate volume <40 cm^3	0.26 (0.11–0.41)	29.5 (15.8–43.3)	Reference	Reference
Lesion size ≥median[a]	0.36 (0.20–0.52)	40.6 (27.5–53.7)	0.93 (0.26–3.37)	5.26 (2.50–11.06)
Lesion size <median[a]	0.27 (0.15–0.38)	29.8 (18.3–41.3)	Reference	Reference
≥4 YoE	0.31 (0.22–0.41)	34.3 (25.7–43.0)	0.57 (0.22–1.46)	0.75 (0.39–1.46)
<4 YoE	0.31 (0.22–0.40)	35.0 (26.6–43.3)	Reference	Reference
Active surveillance	0.36 (0.17–0.55)	39.7 (23.5–55.9)	1.55 (0.47–5.09)	1.37 (0.66–2.87)
Prior negative biopsy	0.42 (0.09–0.76)	46.0 (16.0–76.0)	0.70 (0.08–6.01)	1.03 (0.30–3.60)
Biopsy-naïve	0.26 (0.15–0.37)	29.3 (18.8–39.8)	Reference	Reference

GG = prostate cancer grade group; OR = odds ratio for reader detection of cancer in a positive case given the indicated condition; PPA = positive percent agreement; PSA = prostate-specific antigen; PSAD = PSA density; YoE = years of experience in reading micro-ultrasound.
[a] Owing to angular sagittal frame acquisition, lesion volumes in voxels are not true three-dimensional volumes.

used digital annotation to allow precise volumetric analysis of lesion overlap. In addition to biopsy results, we used 3D co-registration with MRI rather than a cognitive approach to increase the accuracy of reference lesions on MUS. Finally, this is the first study of MUS performance in a cohort in which the prevalence of anterior lesions matched that in the general PCa population. Owing to lower performance in this anatomic region, existing reports on MUS performance should be interpreted with consideration of the zonal cancer distribution in the relevant study cohort.

On the basis of these findings, MUS performance does not yet approach that of MRI in guiding targeted biopsy. Both inter-reader agreement and sensitivity appear to be lower than results reported for MRI. However, the most experienced urologists have <10 yr of experience with MUS. In comparison to MRI, which took uroradiologists >15 yr of research to master both acquisition and interpretation, MUS is in its infancy. While we await the validation of better protocols for identifying lesions in the anterior prostate, it may be valuable to add additional systematic sampling of the anterior prostate to close the gap between imaging modalities. The next steps to improve MUS performance could also include the development of automated cancer detection on live MUS images via artificial intelligence frameworks, which have already shown significant promise in prostate MRI [30]. Such approaches would be particularly advantageous in improving inter-reader agreement.

5. Conclusions

Inter-reader agreement on MUS biopsy targets is lower than what has been reported for MRI. While MUS sensitivity in the posterior prostate approaches previously published rates, performance in the anterior gland is lacking. Further work is needed to determine whether the gap between MUS and MRI in guiding targeted biopsy can be closed.

Author contributions: Steve R. Zhou had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Zhou, Kinnaird, Brisbane, Rusu, Sonn.
Acquisition of data: Zhou, Choi, Kinnaird, Brisbane, Lughezzani, Maffei, Fasulo, Albers, Kornberg, Shao, Sonn.
Analysis and interpretation of data: Zhou, Choi, Vesal, Zhang, Rusu, Sonn.
Drafting of the manuscript: Zhou.
Critical revision of the manuscript for important intellectual content: Zhou, Choi, Vesal, Kinnaird, Brisbane, Lughezzani, Shao, Rusu, Sonn.
Statistical analysis: Zhou.
Obtaining funding: None.
Administrative, technical, or material support: Zhou, Choi, Vesal, Zhang, Fan, Shao.
Supervision: Rusu, Sonn.
Other: None.

Financial disclosures: Steve R. Zhou certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: Geoffrey A. Sonn is a consultant for miR Scientific and Sonablate. Giovanni Lughezzani was a paid consultant for ExactVu in 2021. The remaining authors have nothing to disclose.

Funding/Support and role of the sponsor: None.

References

[1] Hamdy FC, Donovan JL, Lane JA, et al. 10-Year outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer. N

Engl J Med 2016;375:1415–24. https://doi.org/10.1056/NEJMoa1606220.
[2] Pierorazio PM, Walsh PC, Partin AW, Epstein JI. Prognostic Gleason grade grouping: data based on the modified Gleason scoring system. BJU Int 2013;111:753–60. https://doi.org/10.1111/j.1464-410X.2012.11611.x.
[3] Elkhoury FF, Felker ER, Kwan L, et al. Comparison of targeted vs systematic prostate biopsy in men who are biopsy naive. JAMA Surg 2019;154:811. https://doi.org/10.1001/jamasurg.2019.1734.
[4] Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-targeted or standard biopsy for prostate-cancer diagnosis. N Engl J Med 2018;378:1767–77. https://doi.org/10.1056/NEJMoa1801993.
[5] Faria R, Soares MO, Spackman E, et al. Optimising the diagnosis of prostate cancer in the era of multiparametric magnetic resonance imaging: a cost-effectiveness analysis based on the Prostate MR Imaging Study (PROMIS). Eur Urol 2018;73:23–30. https://doi.org/10.1016/j.eururo.2017.08.018.
[6] Manley BJ, Brockman JA, Raup VT, Fowler KJ, Andriole GL. Prostate MRI: a national survey of urologist's attitudes and perceptions. Int Braz J Urol 2016;42:464–71. https://doi.org/10.1590/S1677-5538.IBJU.2015.0235.
[7] Sosnowski R, Zagrodzka M, Borkowski T. The limitations of multiparametric magnetic resonance imaging also must be borne in mind. Cent Eur J Urol 2016;69:22–3. https://doi.org/10.5173/ceju.2016.e113.
[8] Rohrbach D, Wodlinger B, Wen J, Mamou J, Feleppa E. High-frequency quantitative ultrasound for imaging prostate cancer using a novel micro-ultrasound scanner. Ultrasound Med Biol 2018;44:1341–54. https://doi.org/10.1016/j.ultrasmedbio.2018.02.014.
[9] Pedraza AM, Gupta R, Musheyev D, et al. Microultrasound in the detection of the index lesion in prostate cancer. Prostate 2024;84:79–86. https://doi.org/10.1002/pros.24628.
[10] Albers P, Wang B, Broomfield S, Medina Martín A, Fung C, Kinnaird A. Micro-ultrasound versus magnetic resonance imaging in prostate cancer active surveillance. Eur Urol Open Sci 2022;46:33–5. https://doi.org/10.1016/j.euros.2022.09.019.
[11] Sountoulides P, Pyrgidis N, Polyzos SA, et al. Micro-ultrasound-guided vs multiparametric magnetic resonance imaging-targeted biopsy in the detection of prostate cancer: a systematic review and meta-analysis. J Urol 2021;205:1254–62. https://doi.org/10.1097/JU.0000000000001639.
[12] Brembilla G, Dell'Oglio P, Stabile A, et al. Interreader variability in prostate MRI reporting using Prostate Imaging Reporting and Data System version 2.1. Eur Radiol 2020;30:3383–92. https://doi.org/10.1007/s00330-019-06654-2.
[13] Greer MD, Shih JH, Lay N, et al. Interreader variability of Prostate Imaging Reporting and Data System version 2 in detecting and assessing prostate cancer lesions at prostate MRI. Am J Roentgenol 2019;212:1197–205. https://doi.org/10.2214/AJR.18.20536.
[14] Kohestani K, Wallström J, Dehlfors N, et al. Performance and inter-observer variability of prostate MRI (PI-RADS version 2) outside high-volume centres. Scand J Urol 2019;53:304–11. https://doi.org/10.1080/21681805.2019.1675757.
[15] Turkbey B, Rosenkrantz AB, Haider MA, et al. Prostate Imaging-Reporting and Data System version 2.1: 2019 update of Prostate Imaging-Reporting and Data System version 2. Eur Urol 2019;76:340–51. https://doi.org/10.1016/j.eururo.2019.02.033.
[16] Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging 2012;30:1323–41. https://doi.org/10.1016/j.mri.2012.05.001.
[17] Ghai S, Eure G, Fradet V, et al. Assessing cancer risk on novel 29 MHz micro-ultrasound images of the prostate: creation of the micro-ultrasound protocol for prostate risk identification. J Urol 2016;196:562–9. https://doi.org/10.1016/j.juro.2015.12.093.
[18] Light RJ. Measures of response agreement for qualitative data: some generalizations and alternatives. Psychol Bull 1971;76:365–77. https://doi.org/10.1037/h0031643.
[19] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37–46. https://doi.org/10.1177/001316446002000104.
[20] Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull 1971;76:378–82.
[21] Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 2012;8:23–34. https://doi.org/10.20982/tqmp.08.1.p023.
[22] Shih JH, Greer MD, Turkbey B. 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 2018;25:1325–32. https://doi.org/10.1016/j.acra.2018.01.030.
[23] Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. J Clin Epidemiol 1990;43:551–8. https://doi.org/10.1016/0895-4356(90)90159-m.
[24] Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of kappa and weighted kappa. Psychol Bull 1969;72:323–7. https://doi.org/10.1037/h0028106.
[25] Graham P, Bull B. Approximate standard errors and confidence intervals for indices of positive and negative agreement. J Clin Epidemiol 1998;51:763–71. https://doi.org/10.1016/S0895-4356(98)00048-1.
[26] Harland N, Stenzl A. Micro-ultrasound: a way to bring imaging for prostate cancer back to urology. Prostate Int 2021;9:61–5. https://doi.org/10.1016/j.prnil.2020.12.002.
[27] Schaer S, Rakauskas A, Dagher J, et al. Assessing cancer risk in the anterior part of the prostate using micro-ultrasound: validation of a novel distinct protocol. World J Urol 2023;41:3325–31. https://doi.org/10.1007/s00345-023-04591-w.
[28] McNeal JE, Redwine EA, Freiha FS, Stamey TA. Zonal distribution of prostatic adenocarcinoma. Am J Surg Pathol 1988;12:897–906. https://doi.org/10.1097/00000478-198812000-00001.
[29] Johnson DC, Raman SS, Mirak SA, et al. Detection of individual prostate cancer foci via multiparametric magnetic resonance imaging. Eur Urol 2019;75:712–20. https://doi.org/10.1016/j.eururo.2018.11.031.
[30] Bhattacharya I, Seetharaman A, Kunder C, et al. Selective identification and localization of indolent and aggressive prostate cancers via CorrSigNIA: an MRI-pathology correlation and deep learning framework. Med Image Anal 2022;75:102288. https://doi.org/10.1016/j.media.2021.102288.