Can the positive predictive value of prostate MRI in correlation with biopsy findings be interpreted without diving into details?

Supporting material

Ivan Jambor[1,2], Ugo Giovanni Falagario[3], Alberto Martini[4], David Eldred-Evans[5], Hashim U. Ahmed[5], Peter J. Boström[6]

On behalf of the PRostate Mri Outcome Database (PROMOD) study group

1 Department of Radiology, Icahn School of Medicine at Mount Sinai, New York, USA

2 Department of Radiology, University of Turku, Turku, Finland

3 Department of Urology and Organ Transplantation, University of Foggia, Foggia, Italy

4 Department of Oncology/Unit of Urology, Urological Research Institute, IRCCS Ospedale San Raffaele, Milan, Italy.

5 Imperial Prostate, Division of Surgery, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, UK

6 Department of Urology, University of Turku and Turku University hospital, Turku, Finland;

1. Patient selection: The authors present PPV stratified only based on first, PIRADSv.2.0 (PI-RADSv2.0 score 1-2 vs. 3-5 and PI-RADs 1-3 versus 4-5) and prior biopsy status (Westphalen et al. Table E4). Since PPV is higher if the prevalence of the disease is high, a more detailed description of the clinical features of the patients by each center is required. Patient stratification based on the history of prostatic biopsies only partially mitigates the differences in disease prevalence in each center. In addition, details about the type of prior biopsy, when done, is relevant considering targeting, systematic only and saturation biopsy strategies will lead to differing disease prevalence. It is known that patients with a negative targeted biopsy have a lower probability of having a second positive targeted biopsy. Similarly, patients with a previous cancer diagnosis by means of a targeted biopsy, rather than a standard biopsy, are more likely to have a second positive targeted biopsy.

The authors tried to account for differences in prevalence of clinically significant PCa [(csPCa), *i.e.* GG>/=2,] between centers by fitting a mixed-model logistic regression. However, we believe that such an analysis may be unreliable as it did not incorporate a number of factors impacting on PPV, such as prostate volume and/or use of 5-alpha-reductase inhibitors. Race, family history of PCa, PSA, digital rectal examination findings, are the only patient cohort variables collected in the RedCap database presented by the authors.

2. MRI acquisition and reconstruction protocols: No details were provided regarding MRI acquisition and reconstruction protocol. We presume the figures were selected as good examples of MRI but even in those snapshots obvious artefacts can be seen. Despite the authors stating "MRI scans were acquired and interpreted according to the PI-RADS recommendations, followed by MRI-targeted biopsy", without access to the imaging datasets allowing readers to themselves evaluate MRI data quality this cannot necessarily be taken at face value. In fact, only three (T2

planes, DWI max b-value and temporal resolution/duration of DCE) MRI acquisition parameters are presented in TableE1 which do not allow for the evaluation of quality of each MRI acquisition, and reconstruction protocol or implementation of protocols at other centers.

3. MRI reporting system: the authors state that all included MRI studies were reported according to PIRADs version 2 (presumably version 2.0) and results are presented stratified based on lesion location, use of endorectal coil, and magnet strength (Table E5). However, variations in PPV between readers at the same institution reporting MRI datasets collected using the same MRI protocol are not presented. According to the presented RedCap database, MRI readers' indexes per each center as well as for operators performing TB were not collected. Moreover, it is unclear in what form MRI reports were presented to operators performing TB (text, text+images, 3D annotated volumes for fusion, …).

4. Biopsy technique & system: the authors collected in their RedCap dataset if Cognitive/Fusion/In-bore and TB/TB+SB was performed but they did not collect data on the number of biopsy cores for each lesion. We believe this a crucial information and propose that the number of biopsy cores of TB per each lesion should be collected as well as the number and scheme of SB cores in future studies.

5. Histopathology: It remains unclear how the histopathological material was reported. Specifically, how Gleason score was defined for TB versus SB and how Gleason score for each reported lesion was derived. In the presented RedCap data form, Gleason score per lesion was collected but no information on biopsy cores were collected. We believe biopsy core lengths, PCa lengths and Gleason score per core are needed, at least partly, to explore reasons behind wide variations in performance of TB. Further, did pathologists report count intervening benign glands between two foci of cancer within the percent or length of cancer or just the amount of cancer

alone? Ideally, access to digitalized histopathological material would be provided allowing external evaluation of Gleason grading, ideally by semi- or fully automatic methods (1,2). We propose that future studies evaluating performance of prostate MRI in correlation with biopsy findings consider these variables in analyses and ideally provide free public access to digitalize stained histopathological material.

6. Follow up: Authors do not present information if also patient with initially negative targeted biopsy were included. Moreover, no clinical follow up was collected, for example if patients underwent prostatectomy with whole mount prostatectomy sections. Access to such details would allow exploration of further reasons such as biopsy targeting errors.

7. Focusing only on PPV: PPV is a balance of over and under-detection. Thus, PPV (Figure 1) needs to be interpreted in the context of Negative Predicative Value, NPV, (Figure 2). Ideally, sensitivity, specificity, accuracy should be presented since these measures are independent of disease prevalence. The ability to avoid an immediate biopsy is arguably one of the key benefits of an upfront pre-biopsy MRI. However, NPV, sensitivity, specificity, accuracy as well as likelihood ratios can be reliable only if all men receive the same standard test of reference to avoid verification bias. In a study by Khoo CC at al (3) only men with a Likert/PI-RADS score ≥4 or a score of 3 with PSA density ≥0.12 ng/mL/mL were offered biopsy in contrast to a study by Perez IM et al. (4) where men with IMPROD bpMRI Likert score of 1-2 received systematic 12-core biopsy while men with IMPROD bpMRI Likert score ≥3 received targeted biopsy and systematic 12-core biopsy. Thus, direct comparison of NPV, sensitivity, specificity, accuracy and likelihood ratios between these studies cannot be performed and Figure 2 needs to be interpreted in this context.

We believe a thoughtful honest approach with attention to details is needed to allow maximal benefit of prostate MRI to men with suspected or diagnosed PCa. We propose development of a publicly available database which would serve as a benchmark and reference place for centers performing prostate MRI.

Figures Legends

Figure 1 Positive predictive value of different prostate MRI reporting systems in studies by Perez IM et al (4), part A, and Khoo CC et al (3), part B.

Figure 2 Negative predictive value of different prostate MRI reporting systems in studies by Perez IM et al (4), part A, and Khoo CC et al (3), part B.

References

1. Ström P, Kartasalo K, Olsson H, et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. Lancet Oncol. Lancet Publishing Group; 2020;21(2):222–232http://www.thelancet.com/article/S1470204519307387/fulltext. Accessed July 13, 2020.
2. Bulten W, Pinckaers H, van Boven H, et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. Lancet Oncol. Lancet Publishing Group; 2020;21(2):233–241http://www.thelancet.com/article/S1470204519307399/fulltext. Accessed July 13, 2020.
3. Khoo CC, Eldred-Evans D, Jaenicke J, et al. Likert vs. PI-RADS v2: A comparison of two radiological scoring systems for detection of clinically significant prostate cancer. Eur Urol Suppl. 2019;18(1):e1865–e1866.
4. Perez IM, Jambor I, Kauko T, et al. Qualitative and Quantitative Reporting of a Unique Biparametric MRI: Towards Biparametric MRI-Based Nomograms for Prediction of Prostate Biopsy Outcome in Men With a Clinical Suspicion of Prostate Cancer (IMPROD and MULTI-IMPROD Trials). J Magn Reson Imaging. 2019;4–6.

Figure 1

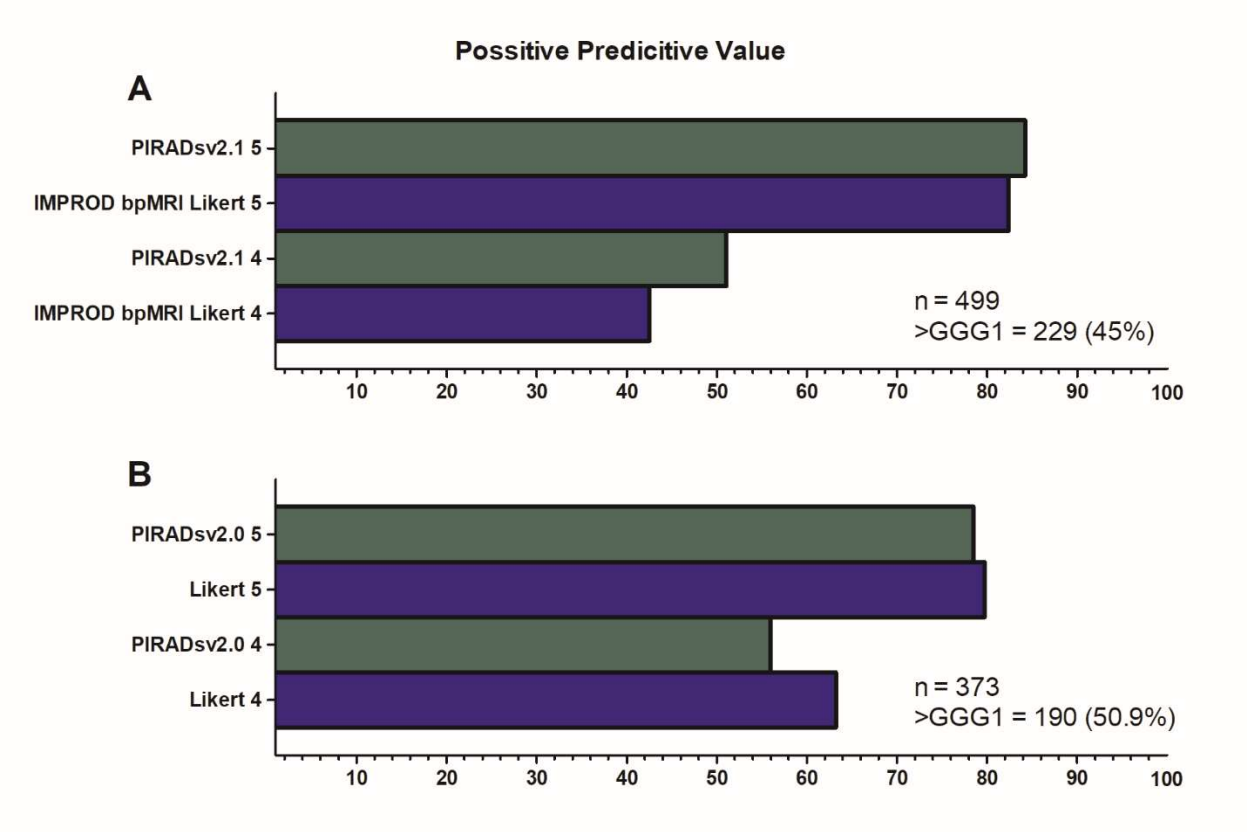
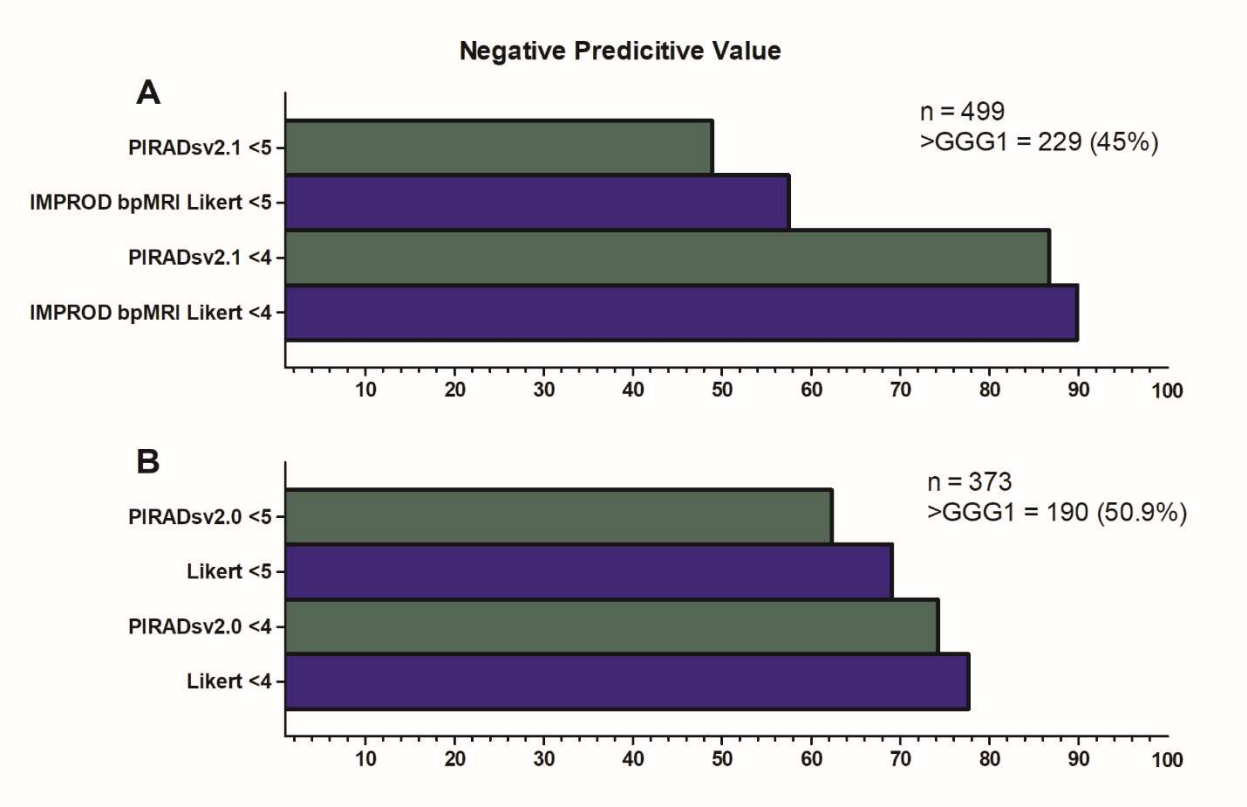

Possitive Predicitive Value
A
PIRADsv2.1 5
IMPROD bpMRI Likert 5
PIRADsv2.1 4
IMPROD bpMRI Likert 4
n = 499
>GGG1 = 229 (45%)
10
20
30
40
50
60
70
80
90
100
B
PIRADsv2.0 5
Likert 5
PIRADsv2.0 4
Likert 4
n = 373
>GGG1 = 190 (50.9%)
10
20
30
40
50
60
70
80
90
100

Figure 2


Negative Predicitive Value
A
n = 499
>GGG1 = 229 (45%)
PIRADsv2.1 <5
IMPROD bpMRI Likert <5
PIRADsv2.1 <4
IMPROD bpMRI Likert <4
10
20
30
40
50
60
70
80
90
100
B
n = 373
>GGG1 = 190 (50.9%)
PIRADsv2.0 <5
Likert <5
PIRADsv2.0 <4
Likert <4
10
20
30
40
50
60
70
80
90
100