



Expected impact of MRI-targeted biopsy interreader variability among urologists on ProScreen prostate cancer screening trial: a pre-trial validation study

Ronja Hietikko^{1,2} · Tuomas Mirtti^{2,3} · Tuomas P. Kilpeläinen^{1,2} · Teemu Tolonen⁴ · Anne Räisänen-Sokolowski^{3,5} · Stig Nordling⁵ · Jill Hannus⁴ · Marita Laurila⁴ · Kimmo Taari¹ · Teuvo L. J. Tammela⁶ · Reija Autio⁷ · Kari Natunen⁷ · Anssi Auvinen⁷ · Antti Rannikko^{1,2}

Received: 8 November 2023 / Accepted: 21 February 2024
© The Author(s) 2024

Abstract

Purpose Prostate cancer (PCa) histology, particularly the Gleason score, is an independent prognostic predictor in PCa. Little is known about the inter-reader variability in grading of targeted prostate biopsy based on magnetic resonance imaging (MRI). The aim of this study was to assess inter-reader variability in Gleason grading of MRI-targeted biopsy among urologists and its potential impact on a population-based randomized PCa screening trial (ProScreen).

Methods From June 2014 to May 2018, 100 men with clinically suspected PCa were retrospectively selected. All men underwent prostate MRI and 86 underwent targeted prostate of the prostate. Six pathologists individually reviewed the pathology slides of the prostate biopsies. The five-tier ISUP (The International Society of Urological Pathology) grade grouping (GG) system was used. Fleiss' weighted kappa (κ) and Model-based kappa for associations were computed to estimate the combined agreement between individual pathologists.

Results GG reporting of targeted prostate was highly consistent among the trial pathologists. Inter-reader agreement for cancer (GG1–5) vs. benign was excellent (Model-based kappa 0.90, Fleiss' kappa $\kappa=0.90$) and for clinically significant prostate cancer (csPCa) (GG2–5 vs. GG0 vs. GG1), it was good (Model-based kappa 0.70, Fleiss' kappa $\kappa 0.67$).

Conclusions Inter-reader agreement in grading of MRI-targeted biopsy was good to excellent, while it was fair to moderate for MRI in the same cohort, as previously shown. Importantly, there was wide consensus by pathologists in assigning the contemporary GG on MRI-targeted biopsy suggesting high reproducibility of pathology reporting in the ProScreen trial.

Keywords Agreement · Interobserver · Grading · Kappa · Pathology

✉ Ronja Hietikko
Ronja.hietikko@hus.fi

- 1 Department of Urology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
- 2 Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Helsinki, Finland
- 3 HUS Diagnostic Center, Department of Pathology, HUS Helsinki University Hospital, Helsinki, Finland
- 4 Fimlab Laboratories, Department of Pathology, Tampere University Hospital, Tampere, Finland
- 5 Department of Pathology, Helsinki University Hospital and University of Helsinki, Helsinki, Finland
- 6 Department of Urology, Tampere University Hospital, Tampere, Finland
- 7 Faculty of Social Sciences, Tampere University, Tampere, Finland

Introduction

Population-based prostate cancer (PCa) screening using prostate-specific antigen (PSA) and standard transrectal ultrasound-guided prostate biopsies in men with elevated PSA levels reduces cancer-specific mortality [1]. However, such screening also results in substantial overdiagnosis and overtreatment of clinically insignificant prostate cancer (cisPCa) [1, 2].

Multiparametric magnetic resonance imaging (mpMRI) of the prostate and subsequent targeted prostate biopsies of identified lesions with clinical suspicion of PCa (PI-RADS 3–5) are a promising diagnostic pathway [3]. In studies involving men with a suspected PCa, mpMRI improves the detection of csPCa and decreases cisPCa diagnosis compared to systematic biopsies [3]. A recent study by Eklund

and colleagues showed that a pre-biopsy MRI only was not inferior to systematic biopsies for detecting csPCa (21% compared to 18%), while detection of cisPCa was reduced by two-thirds[4].

We initiated a population-based, prospective randomized PCa screening trial (ProScreen) in 2018. Unlike the STH-LM3MRI and the Göteborg-2 studies, ProScreen trial is powered to evaluate PCa mortality as the primary endpoint [4, 5]. In the ProScreen trial, screen-positive men are referred to mpMRI with targeted prostate of the MRI visible lesion(s) only [6]. Thus, the emphasis is on minimizing overdiagnosis, while retaining the previously established PCa mortality reduction from screening. To this end, correct identification of csPCa by the pathologists is important for proper treatment selection. Further, to our best knowledge, no previous studies have been published on interobserver agreement of pathologists' interpretation of MRI-targeted prostate biopsies. Importantly, the last ISUP consensus conference emphasized the differences between reporting of systematic biopsies and targeted prostate [7]. Therefore, the aim of this study was to evaluate MRI-targeted biopsy related interreader variability and its expected impact on the ProScreen trial.

Materials and methods

We chose a cohort of 100 men who had been referred to the Helsinki University Hospital (HUS) for suspected PCa before the ProScreen trial. Men had varying baseline risk for PCa. The aim was to evaluate interreader variability in MRI and MRI-targeted biopsy. All 100 men were included in the previously reported study of interreader variability in MRI, and the cohort selection and patient demographics have been reported earlier [8]. For this study, 91 men had undergone MRI before diagnostic biopsies, whereas for 9 men, the MRI was used post-biopsy in cancer staging before definitive treatment. The biopsies were taken between June 2014 and May 2018 using MRI-fusion technique (UroNav, Philips, The Netherlands) to perform transrectal sampling of two to four biopsy cores per suspicious region of interest (ROI). Six patients' samples could not be processed and viewed with cloud viewer due technical issues and were excluded from the final analysis.

All hematoxylin and eosin-stained glass slides of the 85 biopsies were included, representing the full spectrum of Gleason scores and no preselection of any kind was made. Slides were pseudonymized and digitally scanned using Panoramic Flash III slide scanner (3D Histec, Budapest, Hungary) with a pixel resolution of 0.26 $\mu\text{m}/\text{pixel}$ and reviewed with Aiforia cloud viewer software (Aiforia Technologies, Helsinki, Finland).

Six urological pathologists reviewed the slides and filled out a structured pathological assessment query including the number of glasses and biopsies, the length of biopsies and carcinoma, percent of Gleason pattern 4 or 5 and total ISUP Grade Group in each lesion. The pathology reports of the primary ROI (ROI1) were further analyzed. Clinical experience of the pathologists varied from 3 to 50 years (median 12.5, IQR 5.2–35.0). The pathologists were unaware of the other data regarding the patients. The observers were completely independent and were only given instructions to follow the current ISUP guidelines. No formal common training was organized. The original diagnostic pathology report on biopsies of ROI1 was collected.

Statistical analysis

We analyzed the agreement between all pathologists using model-based kappa for association, which is the preferred method when there are more than two raters, and the classifications are ordinal [9]. Model-based kappa for association consider not only the exact agreement, but also the ratings close to each other, and the kappa value is computed giving weights to the classifications. Higher weights are given for the categories that are close to each other [10].

Along Model-based kappa, we have also used Fleiss kappa for comparison. Fleiss kappa is well known and more commonly used, but can be only used for categorical, not ordinal data. Fleiss kappa values and the Model based kappa for associations were computed with R using *irr* and *modelkappa* package. The interobserver agreement for grade groups are illustrated using R package *superheat*. Further the biopsies are clustered with k-means clustering into three groups, while the observers are clustered with hierarchical clustering using Euclidean distance and complete linkage. This retrospective analysis was evaluated and approved by the HUS Research Ethics Committee (HUS/333/2019).

Results

The median age of the study participants at biopsy was 68.8 years (interquartile range [IQR] 60.9–75.0) and the median PSA was 9.1 ng/ml (IQR 6.7–13.8). The median number of biopsies obtained from the index lesion was 2.5 (IQR 2.0–3.0). The reported median length of an individual biopsy was 11.9 mm (IQR 10.4–13.5) and the median length of the cancer in a particular biopsy was 4.9 mm (IQR 3.0–7.8) (Table 1). The distributions of the assigned grades for each case by the study pathologists are shown in Table 2. In the original diagnostic pathology reports, 69 men were diagnosed with PCa. The patient-level GG assessments of all the observers grouped by the original clinical pathology reports are illustrated in Fig. 1.

Table 1 Index lesion-wise comparison of pathological characteristics and agreement on aggregated GG between study pathologists

	Mean Bx length (mm)	Median	SD	Minimum	Maximum	Mean ca length (mm)	Median	SD	Minimum	Maximum	Pathologist experience (years)
Pathologist 1	11.9	11.8	2.5	4.5	23.0	5.9	5.0	3.5	0.5	17.0	15.0
Pathologist 2	12.7	12.5	2.5	6.0	24.0	6.3	6.0	3.6	0.3	17.0	10.0
Pathologist 3	12.1	12.0	2.9	5.0	24.0	5.0	4.5	3.3	0.5	15.0	50.0
Pathologist 4	12.7	12.4	2.6	1.3	23.7	6.1	5.7	3.3	0.4	17.1	3.0
Pathologist 5	12.0	11.9	2.7	5.7	23.7	6.3	5.4	5.7	0.1	17.3	6.0
Pathologist 6	11.0	10.9	2.9	1.7	23.2	5.5	5.1	3.4	0.2	17.2	30.0
						GG1–5 ca			GG3–5 ca		
Agreement ¹ among pathologists 1–6						0.9			0.7		

¹Fleiss' kappa

Table 2 The distribution of the highest ISUP grade 1–5 defined for all patients from index lesion

Proposed grade	Clinical (%)	Pathologist 1 (%)	Pathologist 2 (%)	Pathologist 3 (%)	Pathologist 4 (%)	Pathologist 5 (%)	Pathologist 6 (%)	All pathologists (%)
GG0	20	22	20	20	19	20	18	20
GG1	11	16	22	35	4	31	21	22
GG2	29	15	17	6	20	16	20	16
GG3	21	19	12	8	18	14	11	14
GG4	11	24	15	26	22	11	11	18
GG5	8	4	14	5	18	8	20	11
GG1–5	80	78	80	80	81	80	82	80
GG2–5	69	61	58	45	78	49	61	59
GG3–5	40	46	41	39	58	33	41	43

We found complete agreement on the GG among all (6/6) pathologists in 18 of 85 (21.2%) cases. Of the 18 cases with complete agreement, 72.2% (13/18) were benign. We defined the consensus level as at least 2/3 agreement among the pathologists for a case according to the practice in all consensus meetings organized under the auspices of ISUP during the past decade [11]. The distribution of 2/3 grading consensus for ISUP GGs is shown in Table 3. With this criterion, consensus was reached for 65.9% (56/85) of the cases. The consensus grade differed from the initial grading in 13 cases. Almost all (92.3%) of these cases were in agreement within ±1 of the consensus GG.

No consensus was reached (agreement among pathologists below 2/3) for 34.1% (29/85) of the cases. The most common source of disagreement was the estimated proportion of Gleason patterns 3 and 4. This reflects the challenges in distinguishing GG2 from GG3, as seen in six cases (21.4%).

The agreement among the observers in a comparison including all six categories of cancer and benign was good

(Model-based kappa 0.65, 95% CI 0.59–0.70). Agreement among pathologists for cancer (GG1–5) vs. benign was excellent (Model-based kappa 0.90, Fleiss' kappa $\kappa=0.90$). For three-category comparison between csPCa (GG2–5) vs. cisPCa (GG1) vs. benign (GG0) the interobserver agreement was good (Model-based kappa 0.70, Fleiss' kappa 0.67). The heatmap visualization shows the interobserver agreement for grade groups (Fig. 2).

A total 23 patients had undergone radical prostatectomy until follow-up at the end of June 2019. Pathology report in 35% (n = 8) of these was discordant with the diagnostic pathology report of the MRI-targeted biopsies. Interestingly, and importantly, only 1 of 9 GG2 cancers at biopsy was upgraded to higher grade (one GG2 to GG3) at RP. This is important in the light of MRI-induced grade inflation and supports the notion that GG2 cancers perhaps should increasingly be offered active surveillance instead of immediate curative treatment [12]. However, we also found that the more biopsies were taken, the better concordance was achieved between the pathologist (Table 4).

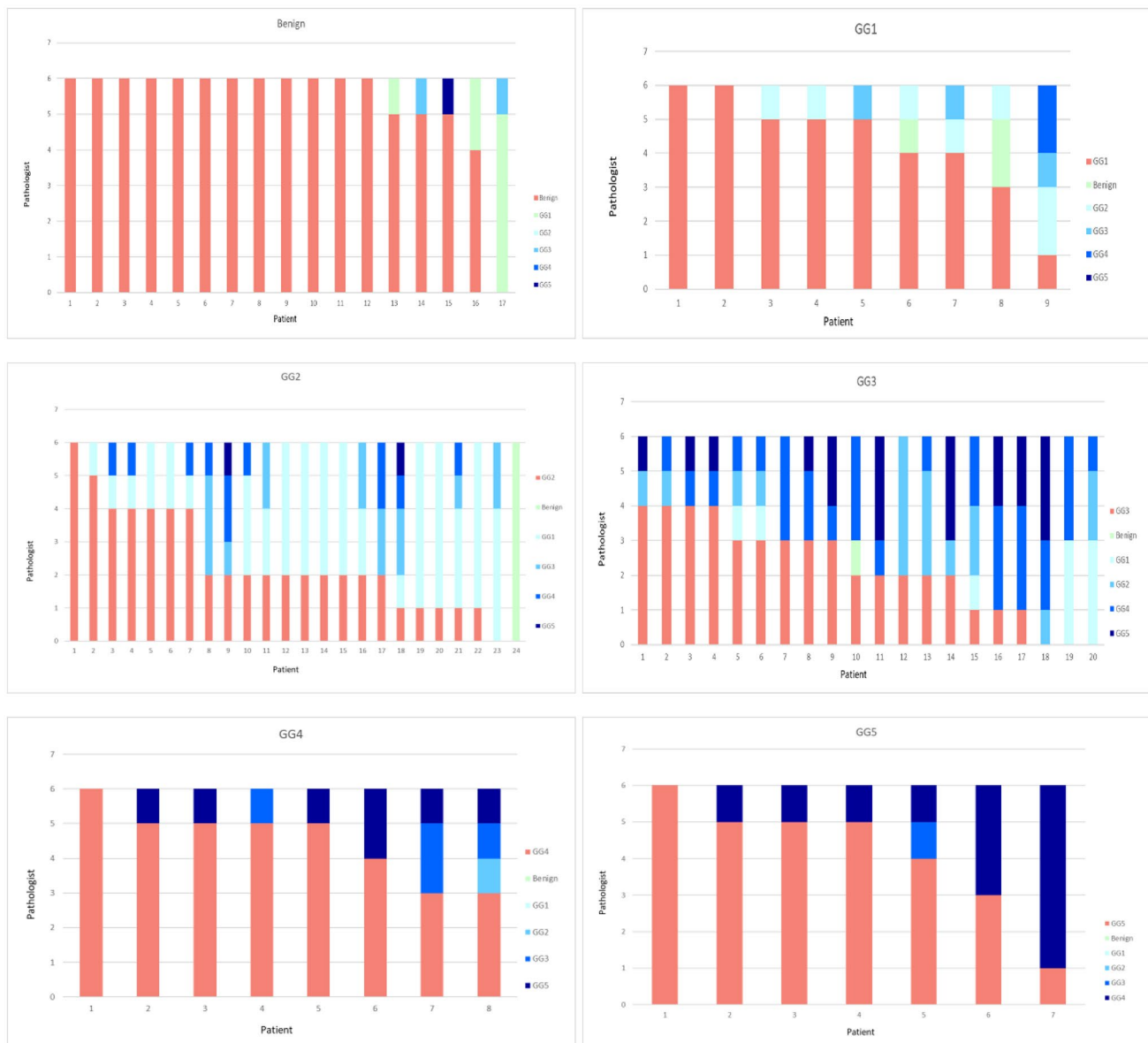


Fig. 1 The number of pathologists identifying Gleason grade in prostate biopsies grouped by original pathological result

Discussion

Despite a recent recommendation by the European Commission to implement PCa in national screening programs, there is no high-quality scientific evidence from randomized screening trials confirming that MRI-based, or a novel biomarker-based, PCa screening would decrease mortality from prostate cancer.

Histological grading is one of the most important prognostic factors of PCa because of its validated prediction of the clinical behavior of cancer [9, 13]. Interestingly, previous screening trials have not assessed the diagnostic agreement of the pathological reporting prior to the study

initiation [1, 14]. Nor have multi-center diagnostic trials comparing MRI-targeted biopsies to systematic biopsies [15, 16]. Here, we show that the interreader agreement among pathologists was good to excellent in grading of MRI-targeted biopsies. Therefore, the expected impact of variability on the MRI-based ProScreen screening trial is minimal. Sufficient agreement between pathologists is crucial for maintaining the value of the Gleason grading system as a diagnostic and prognostic tool and in determining the appropriate treatment for a patient [17]. According to some studies, 10–13% of PCa patients would receive different treatment recommendation after re-evaluation of biopsy specimen [15, 18, 19].

Table 3 Reproducibility by proposed ISUP grade among all cases with the consensus level defined as at least 2/3 of all pathologist

ISUP GG	Consensus <i>n</i> / total <i>n</i>	Consensus for different GG, <i>n</i>	% consensus proposed grade	% consen- sus any grade
0	16/17	1 ^a	94	100
1	7/10	0	70	70
2	7/25	9 ^b	28	64
3	2/17	1 ^c	12	18
4	6/9	1 ^d	67	78
5	5/7	1 ^e	71	86
0–5	56/85	–	–	66

^a An additional case reached consensus for ISUP grade 1
^b Eight additional cases reached consensus for ISUP grade 1, one case reached consensus for grade 0
^c An additional case reached consensus for ISUP grade 2
^d An additional case reached consensus for ISUP grade 3
^e An additional case reached consensus for ISUP grade 4

Table 4 Inter-reader concordance correlated with mean number of biopsies in GG2–3 group

<i>n</i> /6 pathologists agree	Mean number of biopsies
6	4.0
5	3.3
4	2.8
3	3.0
1–2	2.5

In the benign vs. cancer comparison, the more commonly used Fleiss’ kappa was similar to the model-based kappa (Fleiss’ kappa 0.90 vs. Model-based kappa 0.90). The model-based kappa is better suited for multi-categorical association analysis between several observers. Thus, it is not possible to directly compare our results with most of the previously published, systematic biopsy-based studies using Cohen’s or Fleiss’ kappa methodology. In addition, comparison across studies is challenging due to variation in definition of agreement, the type of investigated tissue (e.g., biopsies, radical prostatectomy specimens, transurethral resection specimens, a mixture of these, tissue microarray spots), different grouping of Gleason scores, the number of pathologists involved, and the number of specimens investigated. However, the agreement was better than in most other studies, in which kappa value has been calculated. The reported interobserver agreement among general pathologists for different comparisons has ranged from fair to moderate [20–22] although better results have also been reported [23]. The reproducibility among uropathologists tends to be better than among the general pathologists, usually ranging with between kappa values 0.56–0.67 [20, 23, 24]. We also noticed that the experience of the pathologist influences the results. Observer number four was an outlier in terms of years of experience. When we excluded this pathologist from the analysis, the agreement was higher among the more experienced pathologists.

Similar to our findings on MRI-related inter-reader variation [8], the extremes of the scale seem to be consistently reported, while the intermediate zone with borderline cases is challenging. We found the highest consensus with GG0 (100.0%) and GG5 (85.7%) and lower consensus within GG3 (25.0%) and GG2 (68.0%) cancer. In a PSA-based screening study, as many as 37.5% of biopsies were benign and only 8.0% of diagnosed cancers were GG3, suggesting that the overall reproducibility in a screening may be even better though MRI targeting likely has an impact on this [25]. The most common source of disagreement was separation of Gleason grade pattern 3 from pattern 4. The distinction between these two patterns was also recognized as a challenge in previous studies [23]. Egevad et al. found a specific challenge in differentiating tangentially cut GG1 from GG2 for cases with poorly formed or fused glands [26]. Further,

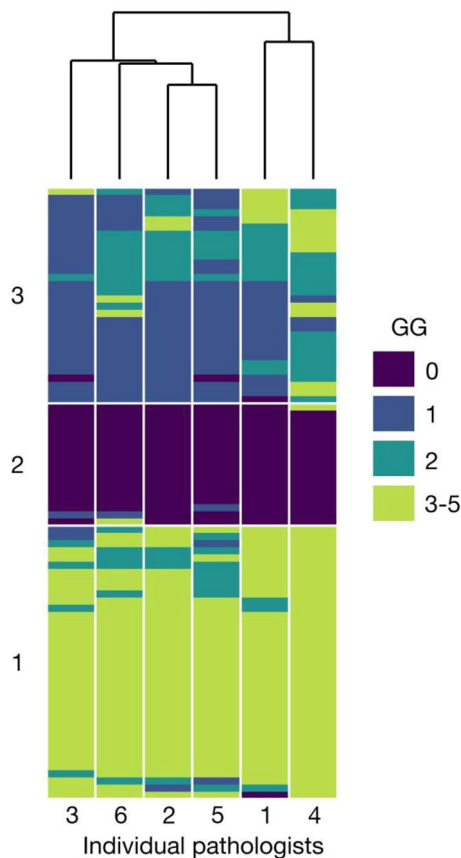


Fig. 2 Heatmap visualization of the interobserver agreement for grade group. Individual pathologist are on the x-axis, colors represent grade groups for each ROI1 biopsy (GG 0=benign). Pathologists and biopsies are ordered based on their similarity resulting from cluster analysis

fused glands or small glands without lumina may be interpreted as tangentially sectioned Gleason grade pattern 3 or as a focal Gleason grade pattern 4 [27]. Zhou et al. reported that any case with ≤ 5 poorly formed glands should not be graded as Gleason pattern 4 [28]. The ISUP 2014 revision of the Gleason grading system suggested that there should be more than occasional structures of this type for a tumor to qualify as Gleason pattern 4, otherwise they may represent tangential cuts. Previous studies have indicated that the reproducibility of Gleason pattern 4 with cribriform pattern is higher than Gleason pattern 4 with poorly formed or fused glands [11, 27]. All the above emphasize the importance of regular training, knowledge exchange between pathologists and intra-institutional peer evaluation, as making the decision on final GG is often subjective, especially in cases composed of Gleason grade patterns 3 and 4. The most evident difference in the fusion biopsy Gleason scoring was in the intermediate GG2/3 group. We suggest that these borderline GG2/3 biopsies should go through a second read. One practical solution for this could be an artificial intelligence-based model, which are available as commercial products and have been shown to improve decision making [29].

As the current clinical practice in diagnosing PCa relies heavily on targeted prostate alone or in conjunction with systematic biopsies, it is important that pathologists follow the guidelines for reporting [7]. Our study is the first to evaluate the interobserver agreement of multiple pathologists on MRI-targeted diagnostic prostate biopsies. Assessing a lesion-wise aggregate GG has been shown to correlate better with RP GG than core-wise highest GG [30], which again emphasizes the need to adhere to reporting guidelines.

The present study has some inherent limitations. MRI-targeted biopsies were obtained from men with clinical suspicion for PCa, not from a screening cohort. This may influence the generalizability of the study results to a screening study with lower underlying PCa risk. However, our study was not designed to assess the diagnostic performance, thus the related limitations such as high prevalence of the disease and verification bias are not essential. Further, the aim was to investigate the interreader agreement among the pathologist for targeted prostate biopsy specifically. Therefore, we chose a study cohort with relatively even distribution of different histopathologies. Moreover, contrary to clinical routine, pathologists were not allowed to consult a colleague when faced with challenging cases. This, however, likely underestimates the interreader agreement observed, especially in more aggressive cancers. When extrapolating the study results on the ProScreen trial, these limitations should not have a major effect as agreement between benign and cisPCa versus csPCa was good, thus supporting clinical decision making on cancer treatments. Given that the same teams of pathologists will evaluate PCa cases in both the screening and control arms, among screening participants

and non-participants, any variability in grading is likely to result in nondifferential misclassification, and hence it is expected to slightly decrease the differences between the compared groups.

Conclusion

The inter-reader agreement of MRI-targeted biopsy was good to excellent and better than the previously published inter-reader agreement for MRI from the same cohort. Therefore, it is plausible to assume that routine clinical histopathological evaluation is not likely to materially impact ProScreen trial results.

Acknowledgements We acknowledge support from the Cancer Foundation Finland and Academy of Finland.

Author contributions RH: protocol development, data management and analysis, manuscript writing and editing. TM: protocol development, data collection and management, manuscript writing and editing. TPK: protocol development, data analysis, manuscript editing. TT: data collection. AR-S: data collection. SN: data collection. JH: data collection. ML: data collection. KT: manuscript revisions. TT: manuscript revisions. RA: data analysis. KN: administrative, technical, and material support. AA: protocol development, manuscript writing and editing. AR: protocol development, manuscript writing and editing.

Funding Open Access funding provided by University of Helsinki (including Helsinki University Central Hospital). This trial was supported by funding from the Cancer Foundation Finland and from Academy of Finland. Author R. Hietikko has received a salary from University of Helsinki.

Data availability The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interests The authors declare no competing interests.

Ethics approval This retrospective analysis was evaluated and approved by the Helsinki University Hospital Research Ethics Committee (HUS/333/2019) and conducted in accordance with Good Clinical Practice guidelines and the Declaration of Helsinki.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Hugosson J, Roobol MJ, Månsson M et al (2019) A 16-years follow-up of the European randomized study of screening for prostate cancer. *Eur Urol* 76(1):43–51. <https://doi.org/10.1016/j.eururo.2019.02.009>
- Welch HG, Albertsen PC (2020) Reconsidering prostate cancer mortality—the future of PSA screening. *N Engl J Med* 382(16):1557–1563. <https://doi.org/10.1056/NEJMms1914228>
- Drost FJH, Osses DF, Nieboer D et al (2019) Prostate MRI, with or without MRI-targeted biopsy, and systematic biopsy for detecting prostate cancer. *Cochrane Database Syst Rev* 4(4):CD012663. <https://doi.org/10.1002/14651858.CD012663.pub2>
- Nordström T, Discacciati A, Bergman M et al (2021) Prostate cancer screening using a combination of risk-prediction, MRI, and targeted prostate biopsies (STHLM3-MRI): a prospective, population-based, randomised, open-label, non-inferiority trial. *Lancet Oncol* 22(9):1240–1249. [https://doi.org/10.1016/S1470-2045\(21\)00348-X](https://doi.org/10.1016/S1470-2045(21)00348-X)
- Kohestani K, Månsson M, Arnsrud Godtman R et al (2021) The GÖTEBORG prostate cancer screening 2 trial: a prospective, randomised, population-based prostate cancer screening trial with prostate-specific antigen testing followed by magnetic resonance imaging of the prostate. *Scand J Urol* 55(2):116–124. <https://doi.org/10.1080/21681805.2021.1881612>
- Auvinen A, Rannikko A, Taari K et al (2017) A randomized trial of early detection of clinically significant prostate cancer (ProScreen): study design and rationale. *Eur J Epidemiol* 32(6):521–527. <https://doi.org/10.1007/s10654-017-0292-5>
- van Leenders GJLH, van der Kwast TH, Grignon DJ et al (2020) The 2019 International society of urological pathology (ISUP) consensus conference on grading of prostatic carcinoma. *Am J Surg Pathol* 44(8):e87–e99. <https://doi.org/10.1097/PAS.0000000000001497>
- Hietikko R, Kilpeläinen TP, Kenttämies A et al (2020) Expected impact of MRI-related interreader variability on ProScreen prostate cancer screening trial: a pre-trial validation study. *Cancer Imaging* 20(1):72. <https://doi.org/10.1186/s40644-020-00351-w>
- Gleason DF (1966) Classification of prostatic carcinomas. *Cancer Chemother Rep* 50(3):125–128
- Nelson KP, Edwards D (2015) Measures of agreement between many raters for ordinal classifications. *Stat Med* 34(23):3116–3132. <https://doi.org/10.1002/sim.6546>
- Egevad L, Delahunt B, Berney DM et al (2018) Utility of pathology imagebase for standardisation of prostate cancer grading. *Histopathology* 73(1):8–18. <https://doi.org/10.1111/his.13471>
- Batouche AO, Czeizler E, Lehto TP et al (2024) MRI-targeted prostate biopsy introduces grade inflation and overtreatment. *medRxiv*. <https://doi.org/10.1101/2024.01.10.24300922>
- 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. *Am J Surg Pathol* 40(2):244–252. <https://doi.org/10.1097/PAS.0000000000000530>
- Pinsky PF, Miller E, Prorok P, Grubb R, Crawford ED, Andriole G (2019) Extended follow-up for prostate cancer incidence and mortality among participants in the prostate, lung, colorectal and ovarian randomized cancer screening trial. *BJU Int* 123(5):854–860. <https://doi.org/10.1111/bju.14580>
- Kasivisvanathan V, Rannikko AS, Borghi M et al (2018) MRI-targeted or standard biopsy for prostate-cancer diagnosis. *N Engl J Med* 378(19):1767–1777. <https://doi.org/10.1056/NEJMoa1801993>
- Klotz L, Chin J, Black PC et al (2021) Comparison of multiparametric magnetic resonance imaging-targeted biopsy with systematic transrectal ultrasonography biopsy for biopsy-naïve men at risk for prostate cancer. *JAMA Oncol* 7(4):534. <https://doi.org/10.1001/jamaoncol.2020.7589>
- Bryant RJ, Sjöberg DD, Vickers AJ et al (2015) Predicting high-grade cancer at ten-core prostate biopsy using four Kallikrein markers measured in blood in the ProtecT study. *J Natl Cancer Inst* 107(7):95. <https://doi.org/10.1093/jnci/djv095>
- Nam RK, Oliver TK, Vickers AJ et al (2012) Prostate-specific antigen test for prostate cancer screening: American Society of Clinical Oncology Provisional Clinical Opinion. *J Oncol Pract* 8(5):315–317. <https://doi.org/10.1200/JOP.2012.000715>
- Bibbins-Domingo K, Grossman DC, Curry SJ (2017) The US preventive services task force 2017 draft recommendation statement on screening for prostate cancer. *JAMA* 317(19):1949. <https://doi.org/10.1001/jama.2017.4413>
- Engers R (2007) Reproducibility and reliability of tumor grading in urological neoplasms. *World J Urol* 25(6):595–605. <https://doi.org/10.1007/s00345-007-0209-0>
- Ozkan TA, Erucar AT, Cebeci OO, Memik O, Ozcan L, Kuskonmaz I (2016) Interobserver variability in Gleason histological grading of prostate cancer. *Scand J Urol* 50(6):420–424. <https://doi.org/10.1080/21681805.2016.1206619>
- Griffiths DFR, Melia J, McWilliam LJ et al (2006) A study of Gleason score interpretation in different groups of UK pathologists; techniques for improving reproducibility. *Histopathology* 48(6):655–662. <https://doi.org/10.1111/j.1365-2559.2006.02394.x>
- Al Nemer AM, Elsharkawy T, Elsharaby M, Al-Tamimi D, Kusaibi H, Ahmed A (2017) The updated grading system of prostate carcinoma: an inter-observer agreement study among general pathologists in an academic practice. *APMIS* 125(11):957–961. <https://doi.org/10.1111/apm.12741>
- Thomsen FB, Marcussen N, Berg KD et al (2015) Repeated biopsies in patients with prostate cancer on active surveillance: clinical implications of interobserver variation in histopathological assessment. *BJU Int* 115(4):599–605. <https://doi.org/10.1111/bju.12820>
- Hugosson J, Månsson M, Wallström J et al (2022) Prostate cancer screening with PSA and MRI followed by targeted biopsy only. *N Engl J Med* 387(23):2126–2137. <https://doi.org/10.1056/NEJMoA2209454>
- Egevad L, Swanberg D, Delahunt B et al (2020) Identification of areas of grading difficulties in prostate cancer and comparison with artificial intelligence assisted grading. *Virchows Arch* 477(6):777–786. <https://doi.org/10.1007/s00428-020-02858-w>
- Kweldam CF, Nieboer D, Algaba F et al (2016) Gleason grade 4 prostate adenocarcinoma patterns: an interobserver agreement study among genitourinary pathologists. *Histopathology* 69(3):441–449. <https://doi.org/10.1111/his.12976>
- Zhou M, Li J, Cheng L et al (2015) Diagnosis of “poorly formed glands” Gleason pattern 4 prostatic adenocarcinoma on needle biopsy. *Am J Surg Pathol* 39(10):1331–1339. <https://doi.org/10.1097/PAS.0000000000000457>
- Baydoun A, Jia AY, Zaorsky NG et al (2024) Artificial intelligence applications in prostate cancer. *Prostate Cancer Prostatic Dis* 27(1):37–45. <https://doi.org/10.1038/s41391-023-00684-0>
- Ren J, Melamed J, Taneja SS et al (2023) Prostate magnetic resonance imaging-targeted biopsy global grade correlates better than highest grade with prostatectomy grade. *Prostate* 83(4):323–330. <https://doi.org/10.1002/pros.24464>

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