BJU Int 2023 doi:10.1111/bju.16163

BJUI
BJU International

Original Article

Performance of magnetic resonance imaging-based prostate cancer risk calculators and decision strategies in two large European medical centres

Petter Davik[1,2], Sebastiaan Remmers[3], Mattijs Elschot[4,5], Monique J. Roobol[3], Tone Frost Bathen[2,4] and Helena Bertilsson[1,2]

[1]*Department of Urology, St Olavs Hospital,* [2]*Department of Clinical and Molecular Medicine (IKOM), Norwegian University of Science and Technology (NTNU), Trondheim, Norway,* [3]*Department of Urology, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam, The Netherlands,* [4]*Department of Radiology and Nuclear Medicine, St Olavs Hospital, and* [5]*Department of Circulation and Medical Imaging (ISB), Norwegian University of Science and Technology (NTNU), Trondheim, Norway*

Patient summary: we compared the decision support tools currently available to help urologists avoid unnecessary prostate biopsies. These tools could help avoid many unnecessary biopsies, whilst missing few cancer cases. The Van Leeuwen model performed the best, followed by the Rotterdam Prostate Cancer Risk Calculator.

Objectives

To compare the performance of currently available biopsy decision support tools incorporating magnetic resonance imaging (MRI) findings in predicting clinically significant prostate cancer (csPCa).

Patients and Methods

We retrospectively included men who underwent prostate MRI and subsequent targeted and/or systematic prostate biopsies in two large European centres. Available decision support tools were identified by a PubMed search. Performance was assessed by calibration, discrimination, decision curve analysis (DCA) and numbers of biopsies avoided vs csPCa cases missed, before and after recalibration, at risk thresholds of 5%–20%.

Results

A total of 940 men were included, 507 (54%) had csPCa. The median (interquartile range) age, prostate-specific antigen (PSA) level, and PSA density (PSAD) were 68 (63–72) years, 9 (7–15) ng/mL, and 0.20 (0.13–0.32) ng/mL^2, respectively. In all, 18 multivariable risk calculators (MRI-RCs) and dichotomous biopsy decision strategies based on MRI findings and PSAD thresholds were assessed. The Van Leeuwen model and the Rotterdam Prostate Cancer Risk Calculator (RPCRC) had the best discriminative ability (area under the receiver operating characteristic curve 0.86) of the MRI-RCs that could be assessed in the whole cohort. DCA showed the highest clinical utility for the Van Leeuwen model, followed by the RPCRC. At the 10% threshold the Van Leeuwen model would avoid 22% of biopsies, missing 1.8% of csPCa, whilst the RPCRC would avoid 20% of biopsies, missing 2.6% of csPCas. These multivariable models outperformed all dichotomous decision strategies based only on MRI-findings and PSAD.

Conclusions

Even in this high-risk cohort, biopsy decision support tools would avoid many prostate biopsies, whilst missing very few csPCa cases. The Van Leeuwen model had the highest clinical utility, followed by the RPCRC. These multivariable MRI-RCs outperformed and should be favoured over decision strategies based only on MRI and PSAD.

Keywords

prostate cancer, prostate cancer diagnosis, magnetic resonance imaging, decision support tools, risk calculator, risk stratification

Introduction

Selection of men for prostate biopsy based on risk of clinically significant prostate cancer (csPCa) is recommended to avoid unnecessary biopsy procedures [1]. MRI of the prostate can, due to its high sensitivity and negative predictive value for csPCa, help reduce biopsy procedures, and in combination with targeted biopsies, increase csPCa detection [2]. Pre-biopsy MRI therefore is recommended, both to avoid unnecessary biopsy procedures and to direct targeted biopsies in positive MRIs (Prostate Imaging-Reporting and Data System [PI-RADS] score ≥ 3) [1,3].

Risk calculators (RCs) use clinical parameters (e.g., PSA, age) to calculate patients' probability of csPCa and thus guide biopsy decisions. Several multivariable RCs that incorporate MRI findings (MRI-RCs) have been developed [4–15], as have various dichotomous PI-RADS and PSA density (PSAD) threshold-based strategies [16]. The MRI-RCs have been shown to outperform MRI as a stand-alone, and non-MRI-RCs, whilst comparisons between MRI-RCs and PI-RADS/PSAD strategies remain scarce [4–8,11,17,18]. The 2022 European Association of Urology (EAU) guidelines recommend using a RC or PI-RADS/PSAD threshold to decide whether to perform biopsy [1], but no consensus exists on which RC or PI-RADS/PSAD threshold should be used.

The aims of this study were to identify the currently published MRI-RCs and MRI-based biopsy decision strategies, to compare the performance of multivariable MRI-RCs with bivariable models based on PI-RADS and PSAD only, and to validate and compare the performance of all MRI-RCs in a large cohort of Dutch and Norwegian men undergoing prostate MRI and biopsy for suspected PCa.

Patients and Methods

Patient Population

A total of 940 consenting patients undergoing prostate MRI and biopsy for suspected PCa were retrospectively registered (Erasmus University Medical Centre, Rotterdam, the Netherlands, n = 518 between 2013 and 2020; St Olavs Hospital, Trondheim, Norway, n = 422 between 2019 and 2022). Patients with previous PCa were excluded. Data from the Rotterdam part of the cohort were previously reported [19]. Patients were not part of previous RC development or validation cohorts. The study was approved by the Erasmus University Medical Centre Institutional Review Board, Rotterdam (METC-2019-0352), and the Regional Ethical Committee, Central Norway (REC-2017/576).

MRI

The MRIs were obtained on a 3-T system without endorectal coil. The imaging protocol included triplane T2-weighted imaging, axial diffusion-weighted and dynamic contrast-enhanced imaging. Reporting was performed by expert uroradiologists using PI-RADS (version 2.0, version 2.1 from 2020) [3].

Biopsy

Patients with PI-RADS ≥ 3 on MRI underwent transperineal or transrectal MRI/ultrasonography image-fusion targeted (± systematic) biopsies. Patients where a clinical suspicion of PCa remained despite a normal MRI (PI-RADS ≤ 2) underwent systematic TRUS-guided biopsies. In all, 69 patients with large peripheral lesions on MRI had TRUS biopsies, as they were deemed unlikely to miss the lesion by the treating urologist. TRUS-guided biopsies were taken in an extended core pattern. Biopsies were taken under local anaesthesia, transrectal with antibiotic prophylaxis and transperineal without. Biopsy specimens were graded by uropathologists according to International Society of Urological Pathology (ISUP) [20], Gleason Score ≥ 3 + 4 (ISUP Grade Group [GG] ≥ 2) considered csPCa.

Model Identification and Construction

A PubMed search was performed for MRI-RCs (Search details in Appendix S1: Supplement—Methods). MRI-RCs predicting csPCa (GG ≥ 2) where model coefficients are published or could be extrapolated from published nomograms, and on-line calculators were included. Corresponding authors of publications with missing model coefficients and/or intercept were contacted. We developed two new logistic regression models based only on PI-RADS and PSAD data from the present cohort. This was to investigate how very parsimonious prediction models compare to multivariable MRI-RCs and to dichotomous decision strategies based only on PI-RADS and PSAD thresholds. The new models were: 'PI-RADS&PSAD category' using PI-RADS and PSAD categories as proposed in previous work by Schoots et al. [16] as the only, categorical input parameter; and 'PI-RADS&PSAD continuous', a bivariable prediction model taking PI-RADS as a categorical and PSAD as a continuous variable (see Appendix S1: Supplement—Methods and Appendix S2: Supplement—Results for details).

Statistical Analyses

Regression formulas were derived from published nomograms and the information available in manuscripts. Missing family history and DRE findings from the Trondheim cohort were imputed by single imputation based on patients' age, PSA, prostate volume, and PI-RADS score ('MICE' package,

version 3.13.0). Probabilities were calculated by RCs where predictors were available. Predictions were compared with biopsy results and calibration assessed by calibration curves, calibration slopes (Ideal 1) and calibration-in-the-large (Ideal 0). Discriminative performance was assessed by the area under the receiver operating characteristic curve (AUC) and clinical utility by decision curve analysis (DCA). DCA provides net benefit, a combined measure of the benefits of a correct classification minus the harms of an incorrect classification by a prediction model at a given threshold [21], shown as net reduction in interventions per 100 patients. Numbers of biopsies saved and csPCa missed using original and recalibrated RCs at pre-defined, clinically relevant biopsy thresholds (5%–20%) were calculated. The RCs were recalibrated to account for differences in csPCa prevalence between RC development cohorts and the present cohort. Recalibration was based on the intercept-in-the-large, where the intercept of each original RC was offset to align the mean of the predicted probabilities with the mean of the observed probabilities. We then compared the clinical utility of recalibrated RCs, as recalibration to the present cohort could have altered the relative performance between RCs. Sensitivity analysis was undertaken by assessing calibration, discriminative ability, and clinical utility for each RC for each study centre and by PI-RADS categories. Statistical analyses were performed in R (R Foundation for Statistical Computing, Vienna, Austria).

Results

Patient Characteristics

Table 1 shows the characteristics for the 940 men included for RC comparison by study centre and for the total cohort. In the total cohort, the median (interquartile range) age, PSA level, prostate volume, and PSAD were 68 (63–72) years, 9 (7–15) ng/mL, 47 (33–68) mL, and 0.20 (0.13–0.32) ng/mL^2, respectively. csPCa was detected in 507 men (54%), 714 (76%) had PI-RADS ≥ 3. DRE was missing for 47 and family history for 90 men in the Trondheim part of the cohort.

Risk Calculators and Strategies Included

Missing model coefficients/intercepts were provided by four corresponding authors. One RC is only applicable to a previous negative biopsy setting [22], whilst two RCs use a different csPCa definition (GG ≥ 3) [23,24], and were therefore excluded. Four dichotomous biopsy decision strategies based on PI-RADS scores, and PSAD thresholds suggested in current guidelines were included [1]. These were: Strategy 1: omit biopsy in PI-RADS 1–3 if PSAD < 0.1 ng/mL; Strategy 2: omit biopsy in PI-RADS 1/2 if PSAD < 0.2 ng/mL and PI-RADS 3 if PSAD < 0.1 ng/mL; Strategy 3: omit biopsy in PI-RADS 1/2 and in PI-RADS 3 if PSAD < 0.15 ng/mL; Strategy 4: omit biopsy in PI-RADS 1/2 and in PI-RADS 3 if

Table 1 Patients' characteristics.

Variable	Cohort		
	Overall (*N* = 940)	**Rotterdam (*N* = 518)**	**Trondheim (*N* = 422)**
Age at visit, years, median (IQR)	68 (63–72)	68 (63–72)	67 (63–72)
PSA level, ng/mL, median (IQR)	9 (7–15)	10 (7–15)	8 (6–14)
Prostate volume, mL, median (IQR)	47 (33–68)	47 (31–69)	48 (35–66)
PSAD, ng/mL^2, median (IQR)	0.20 (0.13–0.32)	0.22 (0.15–0.32)	0.18 (0.12–0.31)
PI-RADS score, *n* (%)			
1/2*	226 (24)	127 (25)	99 (23)
3	147 (16)	92 (18)	55 (13)
4	263 (28)	160 (31)	103 (24)
5	304 (32)	139 (27)	165 (39)
Suspicious DRE, *n* (%)			
0	562 (60)	335 (65)	227 (54)†
1	378 (40)	183 (35)	195 (46)†
Unknown		0	47 (11)
Prior negative biopsy, *n* (%)			
0	578 (61)	217 (42)	361 (86)
1	362 (39)	301 (58)	61 (14)
Family history of PCa, *n* (%)	96 (23)	0 (NA)	96 (23)†
Unknown	608 (65)	518	90 (21)
BMI, kg/m^2, median (IQR)	26.4 (24.7–29.0)	NA (NA)	26.4 (24.7–29.0)
Unknown	518	518	0
csPCa, *n* (%)	507 (54)	243 (47)	264 (63)
Biopsy method, *n* (%)			
Targeted + systematic	510 (54)	275 (53)	235 (56)
Targeted only	139 (15)	116 (22)	23 (5)
Systematic only	291 (31)	127 (25)	164 (39)

**PI-RADS 1 and 2 combined. †After imputation. BMI, body mass index; IQR, interquartile range; NA, not available.*

Table 2 Included RCs with parameters and development characteristics.

RC	Parameters									Development cohort		
	PI-RADS	PSA	PV	Age	DRE	PBx	Fam hx	Race	BMI	*n*	csPCa prev, %	Biopsy method, median cores (*n*)
RPCRC [4]	✓	✓	✓	✓	✓	✓	X	X	X	961	36	TRUS Bx (12) ± TBx if pos. MRI (4)
Van Leeuwen [6]	✓	✓	✓	✓	✓	✓	X	X	X	393	37.90	TPM Bx (30) ± TBx if pos MRI (2)
Mehralivand [5]	✓	✓	✓	✓	✓	✓	X	X	X	400	48.30	TRUS Bx (12) ± TBx if pos. MRI (2)
Radtke [7]	✓	✓	✓	✓	✓	✓	X	X	X	1015	42.00	TPM Bx (24) ± TBx if pos MRI (2)
Barcelona [9]	✓	✓	✓	✓	✓	✓	✓	X	X	1486	36.90	TRUS Bx (12) ± TBx if pos. MRI (2–4)
He [10]	✓	✓	✓	X	X	X	X	X	✓	385	40.50	TPS Bx (12) ± TBx if pos. MRI (3)
Mount Sinai [11]	✓	✓	✓	✓	X	✓	✓	X	X	2363	31.40	TRUS Bx (NA) ± TBx if pos. MRI (NA)
Distler [13]	✓	✓	✓	X	✓	X	X	X	X	1040	43.40	TPS Bx (24) ± TBx if pos. MRI (3)
PLUM [8]	✓	✓	✓	✓	X	✓	X	✓	X	900	30.90	TRUS Bx (6) ± TBx if pos. MRI (NA)
Stanford [14]	✓	✓	✓	✓	X	✓	X	✓	X	1922	46.30	TRUS Bx (NA) ± TBx if pos. MRI (NA)
Bjurlin [12]	✓	✓	✓	✓	X	✓	X	X	X	320	30	TRUS Bx (12) ± TBx if pos. MRI (NA)
Imperial RAPID [15]	✓	✓	✓	✓	X	✓	X	X	X	1189	57	TPS Bx + TBx (min. 3)
PI-RADS + PSAD strategies*	✓	✓	✓	X	X	X	X	X	X	3000†	39	TRUS BX (10–12, 24) ± TBx if pos. MRI (1–2)
PI-RADS + PSAD category	✓	✓	✓	X	X	X	X	X	X	940‡	54	TRUS Bx (12) ± TBx if pos. MRI (4)
PI-RADS + PSAD continuous	✓	✓	✓	X	X	X	X	X	X	940‡	54	TRUS Bx (12) ± TBx if pos. MRI (4)

*BMI, body mass index; Bx, biopsy; csPCa prev., prevalence of clinically significant prostate cancer (GG ≥ 2); Fam hist, family history of prostate cancer; NA, not available; PBx, prior prostate biopsy; pos., positive; PV, prostate volume; TBx, targeted biopsy; TPS, transperineal systematic. *Four dichotomous PI-RADS and PSAD-based biopsy decision strategies. †Metanalysis of five studies (https://bjui-journals.onlinelibrary.wiley.com/doi/10.1111/bju.15277). ‡Present cohort.*

PSAD < 0.2 ng/mL. In addition, the bivariable and categorical models using PI-RADS and PSAD as only predictors were included. Body mass index and family history were not available for Rotterdam patients, three RCs that require these input parameters were therefore evaluated in the Trondheim part of the cohort only [9–11]. The Stanford [8] and Imperial Rapid [15] RCs only apply to MRI-positive men and could therefore only be assessed in 714 patients. In all, 18 MRI-based RCs and decision strategies remained for analysis [4–16], shown in Table 2, with required parameters and development cohort characteristics. The decision support tool identification process is shown in Appendix S2: Supplement—Results.

Calibration and Discrimination

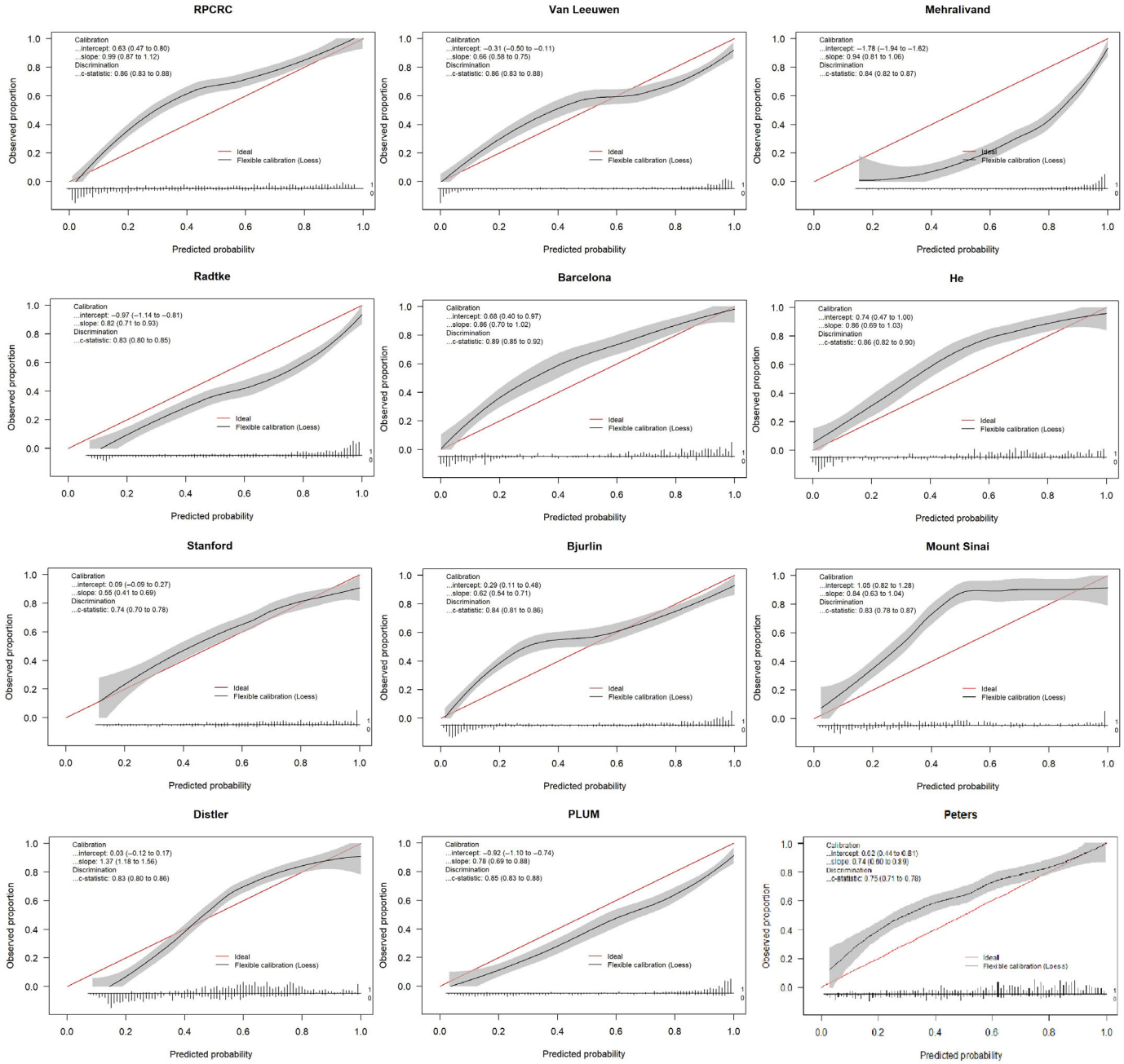
For models where input parameters were available for all 940 patients, the Rotterdam Prostate Cancer Risk Calculator (RPCRC) [4] and the Van Leeuwen model [6] had the best discriminative ability, both with AUC 0.86 (95% CI 0.83–0.88). The Barcelona RC [9] had the highest AUC (0.89, 95% CI 0.85–0.92), but could only be assessed using the Trondheim part of the cohort (n = 422). Figure 1 shows calibration plots before recalibration. The RPCRC [4], Van Leeuwen [6], Prospective Loyola University multiparametric MRI (PLUM) [8], Barcelona [9], He [10], Mount Sinai [11], Bjurlin [12] and Imperial Rapid [15] models were well calibrated at the lower probabilities (≤ 10%) but either underestimated [4,9–12,15], overestimated [8], or gave overly extreme predictions of risks [6,13] at the higher probabilities. The Mehralivand [5], Stanford [8] and Radtke [7] models overestimated risks. The Stanford RC [14] followed the ideal calibration curve line closely, but as for the Imperial Rapid [15], it is only applicable to patients with PI-RADS ≥ 3. Fig. S1 shows calibration plots after recalibration. Table S1 shows discrimination and calibration details for uncalibrated RCs.

Clinical Utility

Decision Curve Analysis

Decision curve analysis of original, uncalibrated RCs (Fig. 2) showed the highest net benefit at all predefined

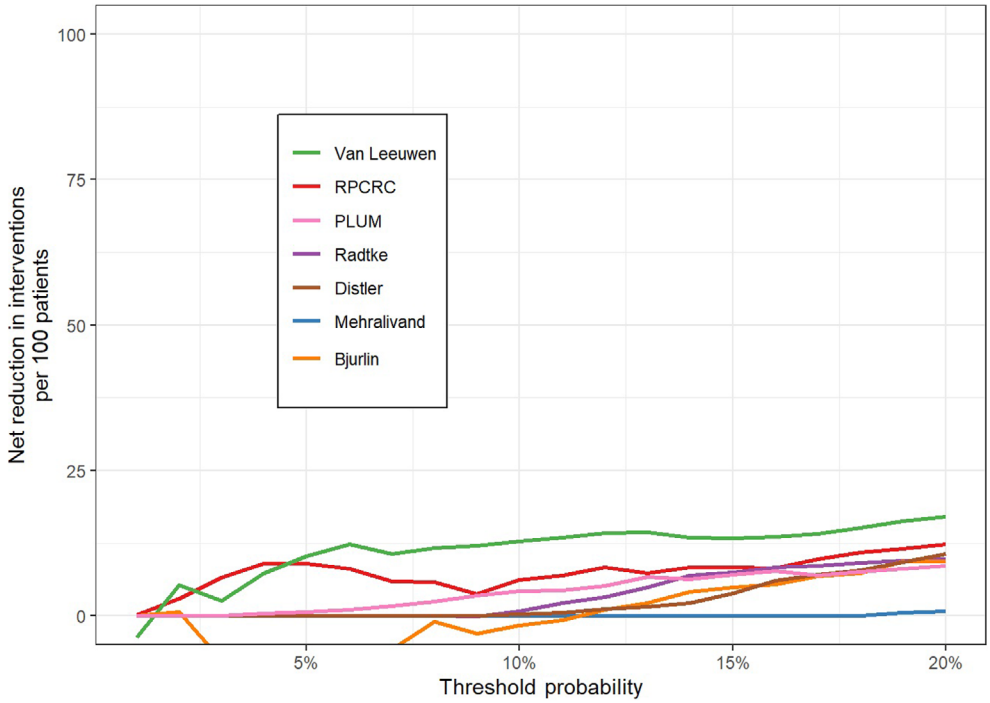
Fig. 1 Calibration plots before recalibration. The calibration slope (Ideal 1) of RPCRC [4] was 0.99 (95% CI 0.87–1.12), followed by the Mehralivand [5] (0.94, 95% CI 0.81–1.06), Barcelona [9], and He [10] RCs (0.86, 95% CI 0.70–1.02; and 0.86, 95% CI 0.69–1.03, respectively). The calibration curve slope of the Van Leeuwen model [6] was 0.66 (95% CI 0.58–0.75). Calibration intercepts ranged between −1.78 and 1.05, with the Distler RC [13] the closest to the ideal 0 (0.03, 95% CI −0.12 to 0.17). Stanford [14] and Imperial Rapid (Peters) [15] RCs only applicable to patients with PI-RADS ≥ 3, He [10], Mount Sinai [11] and Barcelona [9] assessed in Trondheim patients only.



thresholds (5%, 10%, 15% and 20%) for the Van Leeuwen model [6], followed by the RPCRC [4,6]. The PLUM RC provided net benefit at all threshold probabilities, the Radtke and Distler models > 10% threshold probability and the Mehralivand model only at the 20% threshold [5,7,8,13]. The Bjurlin [12] model gave net harm for threshold probabilities ≤ 10%, and net benefit at threshold probabilities 15% and 20%.

Decision curve analysis of recalibrated RCs (Fig. 3) and the present developed 'PI-RADS&PSAD' models showed the highest net benefit for the Van Leeuwen model [6] at

Fig. 2 The DCA (original uncalibrated RCs). The Van Leeuwen [6] model had the highest clinical utility at all biopsy thresholds, followed by the RPCRC [4]. Net reduction in interventions per 100 patients quantified by: (net benefit of model − net benefit biopsy all)/(threshold probability/[1 − threshold probability]).



all thresholds, followed by the RPCRC [4]. The other recalibrated RCs displayed net benefit at all pre-defined thresholds, apart from the Distler model, which did so only at the 15% and 20% thresholds [5,8,13]. The novel PI-RADS&PSAD models (continuous and categorical) gave lower net benefit compared with the best performing MRI-RCs.

Recalibration based on the intercept-in-the-large did not change the main findings of our study.

Biopsies Saved vs csPCa Missed for RCs and PI-RADS and PSAD Decision Strategies

Table 3 shows numbers of biopsy procedures saved and csPCa detected/missed at the pre-defined biopsy thresholds for original, uncalibrated MRI-RCs, and the most closely corresponding PI-RADS/PSAD strategies. Results for recalibrated MRI-RCs are shown in Table S2.

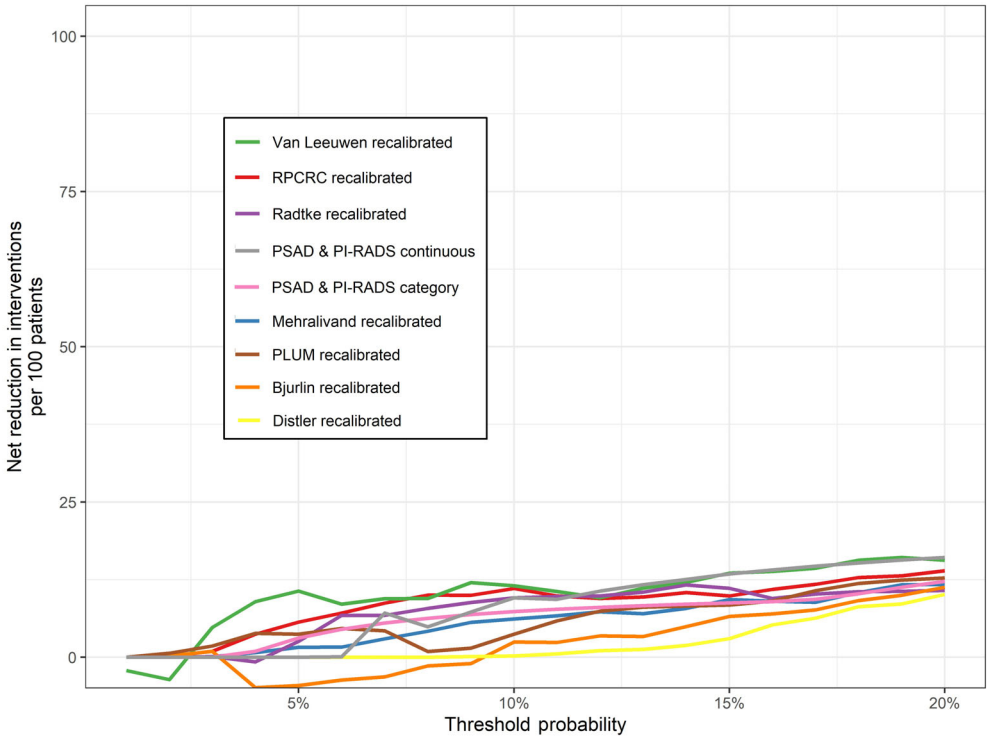
Using the original, uncalibrated Van Leeuwen model [6] would avoid 211 (22%) biopsy procedures at the 10% threshold in our cohort, missing nine (1.8%) cases of csPCa. The original and uncalibrated RPCRC [4] would avoid 188 (20%) of biopsies at the same threshold, missing 13 (2.6%) of csPCa.

Table S3 shows biopsies saved and csPCa detected/missed for each PI-RADS/PSAD biopsy decision strategy, and for each uncalibrated MRI-RC at the threshold that gives the same number of csPCa missed. This to facilitate comparison between dichotomous biopsy decision strategies and RCs that give a probability. All PI-RADS/PSAD strategies avoided fewer biopsies than the best performing MRI-RCs with the same number of missed csPCa. Results for recalibrated MRI-RCs are shown in Table S4.

Sensitivity Analyses

Sensitivity analysis by subgroup analysis for each study centre and by PI-RADS categories did not alter our main study findings. DCA, calibration curves and discrimination details stratified by study centre, are shown in Figs S2–S7. Subgroup analyses of PI-RADS ≤ 3 patients (Figs S8 and S9) showed a maintained ordering of the clinical utility of RCs, with a greater magnitude of net benefit compared to the whole cohort DCA. Subgroup analysis of patients with PI-RADS 3 was inconclusive due to the low number of patients in this group. In subgroup analyses of PI-RADS ≥ 3 patients (Figs S10 and S11) clinical utility was diminished for all included RCs. No patients with PI-RADS 4 or PI-RADS 5 had a calculated probability of ≤ 10% by the Van Leeuwen [6] nor RPCRC models.

Fig. 3 The DCA (recalibrated RCs). PI-RADS&PSAD category—model taking PI-RADS and PSAD categories as the only input parameter. PI-RADS&PSAD continuous—bivariable model taking PI-RADS as a categorical and PSAD as a continuous input parameter. Recalibration based on the intercept-in-the-large improved clinical utility of the Van Leeuwen model [6] at the 5% risk threshold, whilst net benefit diminished or was unchanged at 10%, 15% and 20% thresholds. For the RPCRC [4] recalibration reduced net benefit at the 5% threshold and improved it at the 10%, 15% and 20% thresholds. Recalibration-in-the-large improved clinical utility of the Radtke, Mehralivand, PLUM and Bjurlin models [5,7,8,12] across thresholds, whilst clinical utility of the Distler model [13] was unchanged by recalibration.



Discussion

In this external validation study, we compared the performance of currently available MRI-RCs, several dichotomous PI-RADS/PSAD-threshold-based decision strategies, as well as the present developed PI-RADS/PSAD-based models in a large, contemporary cohort from two large European medical centres. The Van Leeuwen model [6] had the greatest clinical utility, followed by the RPCRC [4], for both the original and recalibrated versions. The best multivariable models outperformed all included dichotomous decision strategies and models based only on MRI-findings and PSAD.

The appropriate risk threshold to recommend biopsy depends on a complete assessment of benefits and harms for a given patient. The original Van Leeuwen [6] and RPCRC publications suggest a 10% biopsy threshold [4,6] to be used. Using the uncalibrated, original Van Leeuwen model [6] this would avoid 22% of biopsies whilst missing 1.8% of csPCa in the present cohort, which is lower than the 28% reduction in biopsies, whilst missing 2.6% of csPCa cases found in the original publication. The uncalibrated RPCRC would give a 20% reduction in biopsies at the same threshold, whilst missing 2.6% of csPCa cases. This is substantially lower than the findings of the original publication where 36% of biopsies were saved and 4% of csPCa missed [4]. This is likely due to the higher csPCa prevalence in the present cohort, compared to the development cohorts (54% vs 36% [RPCRC] and 38% [Van Leeuwen]). At the 4% biopsy threshold suggested on the RPCRC website [25], the uncalibrated RPCRC [4] would avoid 85 biopsies (9%), missing no csPCa, whilst the Van Leeuwen model [6] would avoid 143 biopsies, missing three cases of csPCa (data not shown). Despite the high cancer prevalence in the present cohort, risk-based selection of men for biopsy would reduce unnecessary biopsies substantially at all biopsy risk thresholds, at little risk of missing csPCa.

Some differences are noted between the two best performing MRI-RCs. The Van Leeuwen study [6] applied a more inclusive csPCa definition (Gleason 7 with > 5% GG4, ≥ 20% positive cores or ≥ 7 mm of cancer in any biopsy considered

Table 3 Biopsies saved vs csPCas detected and missed at different risk thresholds of csPCa—uncalibrated RCs.

Risk threshold for csPCa	Men Biopsied, *n* (%)	Biopsies Saved, *n* (%)	csPCas Detected, *n* (%)	csPCas Missed, *n* (%)
5%				
Van Leeuwen	783 (83.3)	156 (16.7)	504 (99.4)	3 (0.6)
RPCRC	835 (88.8)	105 (11.2)	506 (99.8)	1 (0.2)
PLUM	932 (99.4)	6 (0.6)	507 (100)	0 (0)
Mehralivand	940 (100)	0 (0)	507 (100)	0 (0)
Radtke	940 (100)	0 (0)	507 (100)	0 (0)
Distler	940 (100)	0 (0)	507 (100)	0 (0)
Bjurlin	840 (89.4)	100 (10.6)	498 (98.2)	9 (1.8)
Barcelona*	380 (90)	42 (10)	263 (99.6)	1 (0.4)
He*	367 (87)	55 (13)	259 (98.1)	5 (1.9)
Mount Sinai*	415 (98.3)	7 (1.7)	264 (100)	0 (0)
Stanford†	714 (100)	0 (0)	492 (100)	0 (0)
Imperial RAPID†	709 (99.3)	5 (0.7)	792 (100)	0 (0)
PI-RADS&PSAD strategy				
Omit biopsy in PI-RADS 1–3 if PSAD < 0.1 ng/mL^2	879 (93.5)	61 (6.5)	497 (98)	10 (2)
PI-RADS + PSAD classification	831 (88.4)	109 (11.6)	503 (99.2)	4 (0.8)
PI-RADS + PSAD model	940 (100)	0 (0)	507 (100)	0 (0)
10%				
Van Leeuwen	729 (77.6)	211 (22.4)	498 (98.2)	9 (1.8)
RPCRC	752 (80)	188 (20)	494 (97.4)	13 (2.6)
PLUM	900 (95.7)	40 (4.3)	507 (100)	0 (0)
Mehralivand	940 (100)	0 (0)	507 (100)	0 (0)
Radtke	922 (98.1)	18 (1.9)	506 (99.8)	1 (0.2)
Distler	937 (99.7)	3 (0.3)	507 (100)	0 (0)
Bjurlin	725 (77.1)	215 (22.9)	484 (95.5)	23 (4.5)
Barcelona*	347 (82.2)	75 (17.8)	257 (97.3)	7 (2.7)
He*	340 (80.6)	82 (19.4)	255 (96.6)	9 (3.4)
Mount Sinai*	374 (88.6)	48 (11.4)	257 (97.3)	7 (2.7)
Stanford†	714 (100)	0 (0)	492 (100)	0 (0)
Imperial RAPID†	689 (96.5)	25 (3.5)	489 (99.4)	3 (0.6)
PI-RADS&PSAD strategy				
Omit biopsy in PI-RADS 1–3 if PSAD < 0.1 ng/mL^2	879 (93.5)	61 (6.5)	497 (98)	10 (2)
Omit biopsy in PI-RADS 1/2 if PSAD < 0.2 ng/mL^2 and PI-RADS 3 if PSAD < 0.1 ng/mL^2	770 (81.9)	170 (18.1)	493 (97.2)	14 (2.8)
PI-RADS + PSAD classification	835 (88.8)	109 (11.6)	503 (99.2)	4 (0.8)
PI-RADS + PSAD model	720 (76.6)	220 (23.4)	494 (97.4)	13 (2.6)
15%				
Van Leeuwen	701 (74.6)	239 (25.4)	490 (96.6)	17 (3.4)
RPCRC	695 (73.9)	245 (26.1)	482 (95.1)	25 (4.9)
PLUM	853 (90.7)	87 (9.3)	504 (99.4)	3 (0.6)
Mehralivand	940 (100)	0 (0)	507 (100)	0 (0)
Radtke	856 (97.1)	84 (8.9)	505 (99.6)	2 (0.4)
Distler	904 (96.2)	36 (3.8)	507 (100)	0 (0)
Bjurlin	661 (70.3)	279 (29.7)	472 (93.1)	35 (6.9)
Barcelona*	325 (77)	97 (23)	251 (95.1)	13 (4.9)
He*	327 (77.5)	95 (22.5)	252 (95.5)	12 (4.5)
Mount Sinai*	347 (82.2)	75 (17.8)	250 (94.7)	14 (5.3)
Stanford†	705 (98.7)	9 (1.3)	492 (100)	0 (0)
Imperial RAPID†	665 (93.1)	49 (6.9)	481 (97.8)	11 (2.2)
PI-RADS&PSAD strategy				
Omit biopsy in PI-RADS 1/2 if PSAD < 0.2 ng/mL^2 and PI-RADS 3 if PSAD < 0.1 ng/mL^2	770 (81.9)	170 (18.1)	493 (97.2)	14 (2.8)
Omit biopsy in PI-RADS 1/2 and in PI-RADS3 if PSAD < 0.15 ng/mL^2	655 (69.7)	285 (30.3)	473 (93.3)	34 (6.7)
PI-RADS + PSAD classification	835 (88.8)	109 (11.6)	503 (99.2)	4 (0.8)
PI-RADS + PSAD model	714 (76.0)	226 (24.0)	492 (97.0)	15 (3.0)
20%				
Van Leeuwen	680 (72.3)	260 (27.7)	487 (96.1)	20 (3.9)
RPCRC	649 (69)	291 (31)	472 (93.1)	35 (6.9)
PLUM	804 (85.5)	136 (14.5)	496 (97.8)	11 (2.2)
Mehralivand	932 (99.1)	8 (0.9)	507 (100)	0 (0)
Radtke	838 (89.1)	102 (10.9)	505 (99.6)	2 (0.4)
Distler	830 (88.3)	110 (11.7)	505 (99.6)	2 (0.4)
Bjurlin	617 (65.6)	323 (34.4)	460 (90.7)	47 (9.3)
Barcelona*	297 (70.4)	125 (29.6)	243 (92)	21 (8)
He*	319 (75.6)	103 (24.4)	250 (94.7)	14 (5.3)

Table 3 (continued)

Risk threshold for csPCa	Men Biopsied, *n* (%)	Biopsies Saved, *n* (%)	csPCas Detected, *n* (%)	csPCas Missed, *n* (%)
Mount Sinai*	319 (75.6)	103 (24.4)	242 (91.7)	22 (8.3)
Stanford†	685 (95.9)	29 (4.1)	485 (98.6)	7 (1.4)
Imperial RAPID†	642 (89.9)	72 (10.1)	474 (96.3)	18 (3.7)
PI-RADS&PSAD strategy				
Omit biopsy in PI-RADS 1/2 and in PI-RADS 3 if PSAD < 0.15 ng/mL^2	655 (69.7)	285 (30.3)	473 (93.3)	34 (6.7)
Omit biopsy in PI-RADS 1/2 and in PI-RADS 3 if PSAD < 0.2 ng/mL^2	636 (67.7)	304 (32.3)	468 (92.3)	39 (7.7)
PI-RADS + PSAD classification	655 (69.7)	285 (30.3)	473 (93.3)	34 (6.7)
PI-RADS + PSAd model	714 (76.0)	226 (24.0)	492 (97.0)	15 (3.0)

**From patients with complete data, Trondheim cohort (n = 422). †Calculated from patients with positive MRI (PI-RADS > 2) (n = 714).*

as csPCa), and also transperineal mapping biopsies (TPM) which, according to a recent Cochrane report, have a higher sensitivity for detecting csPCa than the MRI-targeted biopsies applied to develop the RPCRC [2,4,6]. These differences could explain the almost identical csPCa proportions in the Van Leeuwen and RPCRC development cohorts despite the lower risk profile of the Van Leeuwen cohort based on rates of positive MRI, PSA, PSAD, and age [4,6]. This could also explain the van Leeuwen model's predictive performance on MRI-targeted biopsies.

Like the Barcelona, Stanford, PLUM, and Mount Sinai RCs, the RPCRC [4,8,9,11,14] is freely accessible on-line, facilitating clinical uptake compared to the Van Leeuwen, Mehralivand, Distler, Radtke, He, and Bjurlin [5–7,10,12,13] models, which are not.

All included RCs showed excellent discriminative ability (AUC 0.8–0.9) apart from the Stanford [14] (AUC 0.74) and Imperial Rapid RCs (AUC 0.75), which are only applicable to MRI-positive men, limiting their performance in the present cohort. Subgroup analysis in MRI-positive men showed improved performance for these two RCs compared with the other MRI-RCs, but only net benefit at the 5% and 10% thresholds for the Imperial Rapid RC [15] and at the 15% threshold for the Stanford RC [14] (Fig. S11). The RPCRC [4] notably had the highest net benefit at the 5% biopsy threshold in MRI-positive men. Calibration and clinical utility differed substantially between MRI-RCs, highlighting the importance of performing external validation in the clinical setting where prediction models are to be used [26]. MRI-RCs that systematically under- or over-estimated risks benefitted the most from recalibration-in-the large [5,7,8,12]. The best performing RCs [4,6] were well calibrated to the present cohort in the clinically relevant probability range (5%–20%) and thus benefitted less from recalibration.

Mortezavi et al. [17] and Saba et al. [27] previously compared four MRI-RCs [4–7] with two non-MRI-RCs and a biomarker test (Mortezavi et al. only) in 532 Scandinavian and 468 Swiss men, respectively. Authors found discriminative abilities like our study (AUCs 0.81–0.87 and 0.73–0.85, respectively). The Van Leeuwen [6] model performed best in the Scandinavian, and the RPCRC [4] in the Swiss study. However, neither study found clinical utility of MRI-RCs at lower risk thresholds (Mortezavi et al. < 10%, Saba et al. < 15% thresholds). Saba et al. [27] applied TPM biopsies with additional cores from suspicious MRI lesions, which could explain some of the difference in performance compared with the present study. Deniffel et al. [18] compared the same four MRI-RCs [4–7] and one PI-RADS/PSAD-based decision strategy (omit biopsy in PI-RADS 3 if PSAD < 0.1 ng/mL^2) in 385 German men. No net benefit was found under the 10% threshold for the Van Leeuwen [6] nor Mehralivand [5] RCs, and MRI-RCs were inferior to the PI-RADS/PSAD strategy up to the 15% threshold. These findings differ from the present study, where net benefit was observed even at the 5% threshold for several uncalibrated and recalibrated MRI-RCs. However, the study by Deniffel et al. [18] included only MRI-positive patients. This gives a selected cohort of patients where the benefit of the risk stratification provided by MRI-RCs is expected to be low and could explain differences to the present study.

Not all input parameters were available for all patients, thus the performance of all MRI-RCs could not be assessed in the full cohort. However, sensitivity analyses did not suggest that overall results would have been different had all input parameters been available from all patients. Other limitations include that analysis was retrospective and only from large European academic centres, therefore results may not be applicable to all clinical settings. We also recognise that long-term outcomes of using MRI-RCs to omit prostate biopsies have not been elucidated, and that this should be addressed in future works.

Despite these limitations, the present work to our knowledge represents the largest and most comprehensive comparison of currently available MRI-based decision support tools and strategies predicting csPCa on biopsy.

Conclusions

We assessed 18 MRI-based RCs and decision strategies for avoiding unnecessary prostate biopsies. The Van Leeuwen model [6] performed the best at all pre-defined probability thresholds. The best multivariable RCs outperformed and should be favoured over all included dichotomous PI-RADS/PSAD-threshold-based strategies. Even in this high-risk cohort, MRI-RCs reduced the number of unnecessary prostate biopsies considerably, at little risk of missing csPCa.

Disclosure of Interests

None declared.

Funding

Research Council of Norway, Grant-Nr: 295013.

Acknowledgements

Maxi Wess, CIMORe, Trondheim programming assistance. Renee Hogenhout, Erasmus, Rotterdam data sharing.

References

1 **Mottet N, van den Bergh RCN, Briers E et al.** EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer-2020 update. Part 1: screening, diagnosis, and local treatment with curative intent. *Eur Urol* 2021; 79: 243–62

2 **Drost FJH, Osses DF, Nieboer D et al.** Prostate MRI, with or without MRI-targeted biopsy, and systematic biopsy for detecting prostate cancer. *Cochrane Database Syst Rev* 2019; 4: CD012663

3 **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

4 **Alberts AR, Roobol MJ, Verbeek JFM et al.** Prediction of high-grade prostate cancer following multiparametric magnetic resonance imaging: improving the Rotterdam European randomized study of screening for prostate cancer risk calculators. *Eur Urol* 2019; 75: 310–8

5 **Mehralivand S, Shih JH, Rais-Bahrami S et al.** A magnetic resonance imaging–based prediction model for prostate biopsy risk stratification. *JAMA Oncol* 2018; 4: 678–85

6 **van Leeuwen PJ, Hayen A, Thompson JE et al.** A multiparametric magnetic resonance imaging-based risk model to determine the risk of significant prostate cancer prior to biopsy. *BJU Int* 2017; 120: 774–81

7 **Radtke JP, Wiesenfarth M, Kesch C et al.** Combined clinical parameters and multiparametric magnetic resonance imaging for advanced risk modeling of prostate cancer—patient-tailored risk stratification can reduce unnecessary biopsies. *Eur Urol* 2017; 72: 888–96

8 **Patel HD, Koehne EL, Shea SM et al.** A prostate biopsy risk calculator based on MRI: development and comparison of the prospective Loyola University multiparametric MRI (PLUM) and Prostate Biopsy Collaborative Group (PBCG) risk calculators. *BJU Int* 2022; 131: 227–35

9 **Morote J, Borque-Fernando A, Triquell M et al.** The Barcelona predictive model of clinically significant prostate cancer. *Cancers* 2022; 14: 1589

10 **He B-M, Shi Z-K, Li H-S et al.** A novel prediction tool based on multiparametric magnetic resonance imaging to determine the biopsy strategy for clinically significant prostate cancer in patients with PSA levels less than 50 ng/ml. *Ann Surg Oncol* 2020; 27: 1284–95

11 **Parekh S, Ratnani P, Falagario U et al.** The Mount Sinai Prebiopsy Risk Calculator for predicting any prostate cancer and clinically significant prostate cancer: development of a risk predictive tool and validation with advanced neural networking, prostate magnetic resonance imaging outcome database, and European randomized study of screening for prostate cancer risk calculator. *Eur Urol Open Sci* 2022; 41: 45–54

12 **Bjurlin MA, Rosenkrantz AB, Sarkar S et al.** Prediction of prostate cancer risk among men undergoing combined MRI-targeted and systematic biopsy using novel pre-biopsy nomograms that incorporate MRI findings. *Urology* 2018; 112: 112–20

13 **Distler FA, Radtke JP, Bonekamp D et al.** The value of PSA density in combination with PI-RADS (TM) for the accuracy of prostate cancer prediction. *J Urol* 2017; 198: 575–82

14 **Wang NN, Zhou SR, Chen L et al.** The Stanford prostate cancer calculator: development and external validation of online nomograms incorporating PIRADS scores to predict clinically significant prostate cancer. *Urol Oncol* 2021; 39: 831.e19–831.e27

15 **Peters M, Eldred-Evans D, Kurver P et al.** Predicting the need for biopsy to detect clinically significant prostate cancer in patients with a magnetic resonance imaging-detected prostate imaging reporting and data system/Likert ≥3 lesion: development and multinational external validation of the imperial rapid access to prostate imaging and diagnosis risk score. *Eur Urol* 2022; 82: 559–68

16 **Schoots IG, Padhani AR.** Risk-adapted biopsy decision based on prostate magnetic resonance imaging and prostate-specific antigen density for enhanced biopsy avoidance in first prostate cancer diagnostic evaluation. *BJU Int* 2021; 127: 175–8

17 **Mortezavi A, Palsdottir T, Eklund M et al.** Head-to-head comparison of conventional, and image- and biomarker-based prostate cancer risk calculators. *Eur Urol Focus* 2021; 7: 546–53

18 **Deniffel D, Healy GM, Dong X et al.** Avoiding unnecessary biopsy: MRI-based risk models versus a PI-RADS and PSA density strategy for clinically significant prostate cancer. *Radiology* 2021; 300: 369–79

19 **Hogenhout R, Remmers S, van Leenders G, Roobol MJ.** The transition from transrectal to transperineal prostate biopsy without antibiotic prophylaxis: cancer detection rates and complication rates. *Prostate Cancer Prostatic Dis* 2023; 26: 581–7

20 **Epstein JI, Allsbrook WC Jr, Amin MB, Egevad LL.** The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of prostatic carcinoma. *Am J Surg Pathol* 2005; 29: 1228–42

21 **Vickers AJ, Elkin EB.** Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006; 26: 565–74

22 **Truong M, Wang B, Gordetsky JB et al.** Multi-institutional nomogram predicting benign prostate pathology on magnetic resonance/ultrasound fusion biopsy in men with a prior negative 12-core systematic biopsy. *Cancer* 2018; 124: 278–85

23 **Fang D, Zhao C, Ren D et al.** Could magnetic resonance imaging help to identify the presence of prostate cancer before initial biopsy? The development of nomogram predicting the outcomes of prostate biopsy in the Chinese population. *Ann Surg Oncol* 2016; 23: 4284–92

24 **Zhang Y, Zeng N, Zhu YC, Huang YXR, Guo Q, Tian Y.** Development and internal validation of PI-RADs v2-based model for clinically significant prostate cancer. *World J Surg Oncol* 2018; 16: 102

25 **Foundation PCR.** SWOP Prostate Cancer Research Foundation. https://www.prostatecancer-riskcalculator.com/. Accessed 14 Feb 2023.

26 **Van Calster B, McLernon DJ, van Smeden M et al.** Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019; 17: 230

27 **Saba K, Wettstein MS, Lieger L et al.** External validation and comparison of prostate cancer risk calculators incorporating multiparametric magnetic resonance imaging for prediction of clinically significant prostate cancer. *J Urol* 2020; 203: 719–26

Correspondence: Petter Davik, St Olavs Hospital, Department of Urology, Trondheim, Norway.

e-mail: petter.davik@ntnu.no

Abbreviations: AUC, area under the receiver operating characteristic curve; DCA, decision curve analysis; GG, Grade Group; ISUP, International Society of Urological Pathology; (cs)PCa, (clinically significant) prostate cancer; PI-RADS, Prostate Image-Reporting and Data Scheme; PLUM, Prospective Loyola University multiparametric MRI; PSAD, PSA density; RC, risk calculator; RPCRC, Rotterdam Prostate Cancer Risk Calculator; TPM, transperineal mapping biopsies.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Supplement—Methods

Appendix S2 Supplement—Results

Fig. S1 Calibration plots after recalibration.

Fig. S2 Calibration Rotterdam.

Fig. S3 Calibration Trondheim.

Fig. S4 The DCA uncalibrated RC Rotterdam.

Fig. S5 The DCA recalibrated RC Rotterdam.

Fig. S6: Uncalibrated RCs Trondheim.

Fig. S7 Recalibrated Trondheim.

Fig. S8 Calibration PI-RADS 1, 2, 3.

Fig. S9 The DCA PI-RADS 1, 2, 3.

Fig. S10 Calibration PI-RADS 3, 4, 5.

Fig. S11 The DCA PI-RADS 3, 4, 5.

Table S1 The MRI RC calibration and discrimination.

Table S2 Biopsies saved vs clinically significant prostate cancers detected using at different biopsy thresholds—recalibrated RCs.

Table S3 Saved and missed for uncalibrated MRI-RCs at rate of missed csPCa by biopsy decision strategy.

Table S4 Saved and missed for recalibrated MRI-RCs at rate of missed csPCa by biopsy decision strategy.