


**UROGENITAL**

**Open Access**



# Diagnostic accuracy of a 'stage-gated' approach for reporting prostate screening MRI: "Is less more?"

Natasha Thorley<sup>1,2</sup>, Tom Parry<sup>1</sup>, Giorgio Brembilla<sup>1,3,4</sup>, Francesco Giganti<sup>2,5</sup>, Tristan Barrett<sup>6</sup>, David Eldred-Evans<sup>7,8</sup>, Nikhil Mayor<sup>7,8</sup>, Alistair Lamb<sup>1</sup>, Penny L. Hubbard Cristinacce<sup>1</sup>, Fiona Gong<sup>1</sup>, Henry H. Tam<sup>9</sup>, Heminder K. Sokhi<sup>10,11</sup>, Anwar R. Padhani<sup>11</sup>, Caroline M. Moore<sup>5,12</sup>, David Atkinson<sup>1</sup>, Hashim U. Ahmed<sup>7,8</sup> and Shonit Punwani<sup>1,2\*</sup> 

## Abstract

**Objectives** To evaluate whether a two-step, 'stage-gated' reporting approach could improve the positive predictive value (PPV) of biparametric (bp)MRI for prostate cancer (PCa) screening compared to conventional Likert/PI-RADS scoring.

**Materials and methods** This retrospective secondary analysis utilised data from IP1-PROSTAGRAM—a prospective, population-based study of men aged 50–69 years who underwent PCa screening with bpMRI, ultrasound and prostate-specific antigen (PSA) testing between October 2018 and May 2019 at two centres (NCT03702439). MRI scans from IP1-PROSTAGRAM were retrospectively evaluated using the 'stage-gated' approach: three radiologists independently reviewed limited MRI sequences (axial T2-weighted and b1500 diffusion-weighted images) and classified scans as positive or negative; if positive, the remaining bpMRI images were reviewed and a hypothetical "decision-to-biopsy" made. The PPV of 'stage-gated' reading was compared to PI-RADS and Likert scores  $\geq 4$  from the original IP1-PROSTAGRAM bpMRI reports. The reference standard was IP1-PROSTAGRAM biopsy results with grade group (GG)  $\geq 2$  cancer considered significant.

**Results** Of 408 participants (median age 57 years [IQR 53, 61]), 405 had MRI scans available for secondary analysis. The prevalence of GG  $\geq 2$  cancer was 4% (17/405). The 'stage-gated' reporting approach achieved a PPV of 53% (95% CI: 30, 75; 8/15), compared to 29% (95% CI: 15, 47; 8/28) and 30% (95% CI: 17, 46; 11/37) for Likert and PI-RADS  $\geq 4$  pathways, respectively. The 'stage-gated' approach halved the number of recommended biopsies while maintaining similar cancer detection rates.

**Conclusion** The 'stage-gated' reporting approach, using limited sequences for the initial read, may improve the PPV and benefit-to-harm ratio of MRI-based screening.

## Key Points

**Question** *The PPV of MRI in PCa screening is low, likely because conventional assessment systems are not optimised for low disease prevalence populations.*

**Findings** *A two-step, 'stage-gated' reading approach achieved a PPV of 53% (8/15), compared to 29% (8/28) for Likert and 30% (11/37) for PI-RADS scores  $\geq 4$ .*

**Clinical relevance** *The 'stage-gated' reporting approach, which uses limited sequences for the initial read, may improve the PPV and benefit-to-harm ratio of MRI-based screening by reducing unnecessary biopsies. Prospective evaluation is needed to confirm these findings in real-world screening settings.*

\*Correspondence:

Shonit Punwani

[s.punwani@ucl.ac.uk](mailto:s.punwani@ucl.ac.uk)

Full list of author information is available at the end of the article



© The Author(s) 2026. **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/>.

**Keywords** Prostatic neoplasms, Mass screening, Magnetic resonance imaging, Prostate

### Graphical Abstract

## Diagnostic accuracy of a ‘stage-gated’ approach for reporting prostate screening MRI: “Is less more?”

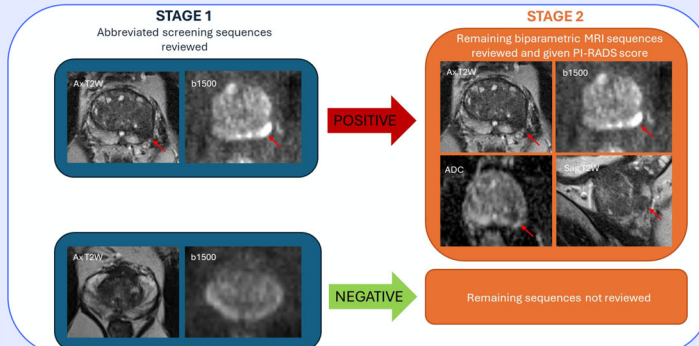
Can a novel two-step reporting approach improve the positive predictive value (PPV) of MRI for prostate cancer screening?

- Biparametric (bp)MRI from 405 IP1-PROSTAGRAM screening trial participants
- MRI re-read with ‘stage-gated’ approach – initial review of limited sequences, with full bpMRI reviewed only if positive
- PPV compared to conventional Likert/PI-RADS scoring ( $\geq 4/5$ ) with biopsy reference standard

Cancer Screening

Prostate

Retrospective, Multi-centre



The ‘stage-gated’ reporting approach, which uses limited sequences for the initial read, improved the PPV and halved the number of recommended biopsies compared to conventional PI-RADS and Likert scoring.

*Eur Radiol* (2025) Thorley N, Parry T, Brembilla G et al;  
DOI: 10.1007/s00330-025-12250-4

EUROPEAN SOCIETY OF RADIOLOGY  
European Radiology

### Introduction

Prostate cancer (PCa) poses a substantial global health burden, and recent European Commission recommendations have urged further research into organised screening programmes [1]. A major challenge is establishing an approach that balances the benefits and harms of screening. Prostate-specific antigen (PSA) testing reduces mortality [2], but has notable limitations, including unnecessary biopsies and overdiagnosis [3, 4], leading to treatment-related harm.

Incorporating MRI into screening strategies may mitigate these harms by reducing unnecessary biopsies and enabling biopsy targeting, potentially improving the detection of clinically significant PCa [5–7]. Emerging evidence suggests shorter, non-contrast biparametric (bp)MRI protocols offer comparable effectiveness to multiparametric (mp)MRI [8–11], while improving feasibility for screening.

While MRI demonstrates high sensitivity for PCa detection, its positive predictive value (PPV) in a screening setting is relatively low. This is likely because conventional assessment systems, such as the Likert scale and prostate imaging reporting and data system (PI-RADS), were

designed for mpMRI in higher-prevalence diagnostic settings and are not optimised for low-prevalence screening populations. For example, the PPV of Likert/PI-RADS scores of  $\geq 4$  on bpMRI in screening contexts ranges from 27% to 46% [12–15], leading to unnecessary biopsies in a significant proportion of men.

In the ReIMAGINE Screening Study [16, 17], abbreviated bpMRI using only axial T2-weighted imaging (T2WI) and high  $b$ -value ( $2000 \text{ s/mm}^2$ ) diffusion-weighted imaging (DWI) at 3-T achieved a PPV of 59% (95% CI: 44, 72), rising to 86% (95% CI: 69, 95) when followed by a separate mpMRI in screen-positive cases [18]. This high PPV suggests that using a limited set of key sequences for the initial read may raise the threshold for positivity and better differentiate between positive and negative scans. While ReIMAGINE involved two separate MRI examinations, with mpMRI at the second stage, we aimed to improve practicality and efficiency for screening by simulating a similar stepwise strategy for reporting a single bpMRI examination. This involved first reviewing limited sequences and proceeding to review the remaining sequences only if the criteria for screen-positivity were met.

In this retrospective study, we utilised the IP1-PROSTAGRAM dataset [12, 19] to evaluate whether this two-step, ‘stage-gated’ reporting approach could improve the PPV of bpMRI for PCa screening compared to conventional Likert/PI-RADS scoring designed for mpMRI reading.

**Materials and methods**

**Study population**

This retrospective secondary analysis used data from the IP1-PROSTAGRAM study (NCT03702439), which received approval from the UK National Research Ethics Committee (8/LO/1338). All participants provided written informed consent. IP1-PROSTAGRAM protocol and outcomes have been previously published [12, 19].

IP1-PROSTAGRAM was a prospective, population-based PCa screening study that recruited men aged 50–69 years between October 2018 and May 2019. Eligible participants had a life expectancy of  $\geq 10$  years and none of the following exclusion criteria: history of PSA testing or prostate MRI within the past 2 years, recent urinary infection or prostatitis, prior prostate biopsy or PCa, or MRI contraindications (Supplementary Table S1).

**IP1-PROSTAGRAM study procedures**

Participants underwent screening with PSA testing, MRI, and ultrasound.

BpMRI sequences included axial and sagittal T2WI, and DWI with multiple *b*-values (0, 150, 400, 1000, and 1500 s/mm<sup>2</sup>) to derive apparent diffusion coefficient (ADC) maps. Acquisition time was 15 min (acquisition parameters are in Supplementary Table S2). Scans were performed at two centres using one 1.5-T (Siemens Magnetom Aera; 17%) and one 3-T scanner (Siemens Magnetom Verio; 83%) with pelvic phased-array coils. Scans with poor image quality were repeated, and participants with susceptibility artefact on DWI due to rectal gas were offered a flatus tube. An antispasmodic agent was administered unless contraindicated.

MRI scans were interpreted by two expert urologists, blinded to PSA and clinical information (apart from age), using PI-RADS v2 [20] (adapted for non-contrast bpMRI) and Likert scores. Ultrasound was independently

scored using a validated 5-point system [21]. A screen-positive result was defined as an MRI or ultrasound score  $\geq 3/5$ , or a PSA  $\geq 3$  ng/mL. A biopsy was recommended if any test was screen positive. Participants underwent transperineal, systematic, 12-core biopsies with additional image fusion-targeted biopsies of all MRI and ultrasound lesions.

**‘Stage-Gated’ study procedures**

In this secondary analysis, MRI scans from 405 participants were consecutively, retrospectively re-evaluated between August 2024 and December 2024 using a ‘stage-gated’ approach (Fig. 1).

**Stage 1**

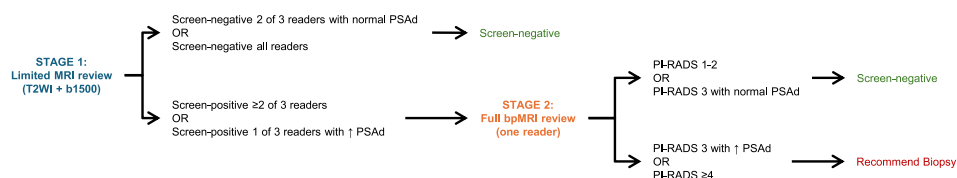
Three expert urologists (G.B., F.G., and T.B., with 8-, 11-, and 15-years’ prostate MRI experience, respectively), who were not involved in the original study, independently reviewed limited MRI sequences comprising only axial T2WI and b1500 DWI. Scans were classified as screen-positive, screen-negative, or non-diagnostic, based on prespecified criteria adapted from ReIMAGINE (Table 1) [18]. In brief, screen-positive scans required abnormal T2WI (score  $\geq 3/5$ ) with corresponding focal abnormality on b1500 DWI (score  $\geq 4/5$ ). Non-diagnostic scans had either (i) both sequences non-diagnostic, or (ii) one non-diagnostic sequence with a suspicious lesion (score  $\geq 4/5$ ) on the other. All other scans were screen-negative. Radiologists were blinded to all clinical and pathological data, including PSA, age, original screening classification, and histopathologic outcomes.

**Advancement to Stage 2**

Scans were advanced to Stage 2 if deemed screen-positive by at least two of three radiologists. In cases where only one radiologist classified a scan as screen-positive, the scan was advanced to Stage 2 if PSA density was elevated (PSAd  $\geq 0.12$  ng/mL<sup>2</sup>).

**Stage 2**

At this stage, the full bpMRI (including ADC maps and sagittal T2WI) was reviewed by one radiologist (G.B., F.G., or T.B.) who assigned a bpMRI-adapted PI-RADS v2.1



**Fig. 1** Schematic representation of the two-step ‘stage-gated’ reporting approach. ↑PSAd raised prostate-specific antigen density ( $\geq 0.12$  ng/mL<sup>2</sup>), T2WI T2-weighted imaging, b1500 b1500 s/mm<sup>2</sup> diffusion weighted imaging

**Table 1** Screening criteria used for stage 1 of the ‘stage-gated’ prostate MRI reporting approach, based on the criteria used in the ReIMAGINE screening study [18]

Axial T2	Axial b1500	Screen outcome
Focal 4–5/5 or diffuse 3/5	Focal 4–5/5	Positive
Focal 4–5/5	Focal 1–2/5 or diffuse 3/5	Negative
Diffuse 3/5	1–2/5 or diffuse 3/5	Negative
1–2/5	ANY	Negative
ANY	1–2/5	Negative
Non-diagnostic (artefact)	1–2/5 or diffuse 3/5	Negative
1–2/5 or diffuse 3/5	Non-diagnostic (artefact)	Negative
4–5/5	Non-diagnostic (artefact)	Non-diagnostic (recall)
Non-diagnostic (artefact)	4–5/5	Non-diagnostic (recall)
Non-diagnostic (artefact)	Non-diagnostic (artefact)	Non-diagnostic (no recall)

Lesions scoring 3/5 on T2 with a focal area that is slightly more intense than background diffuse change (but not qualifying as 4/5) are encompassed under diffuse 3/5

[22] score. The radiologist was blinded to the reason the scan was advanced to Stage 2. A retrospective, hypothetical biopsy recommendation was made for scans scoring PI-RADS  $\geq 4$ , or PI-RADS 3 with elevated PSA<sub>d</sub>.

This stepwise approach balances specificity through reader consensus while preserving sensitivity by incorporating a PSA<sub>d</sub> threshold in discordant (Stage 1) or equivocal (Stage 2) cases. Although three readers reviewed scans concurrently, the outcomes are functionally equivalent to a two-reader workflow with a third reader acting as an arbitrator, and this design enabled assessment of alternative screening pathways. The PSA<sub>d</sub> threshold ( $\geq 0.12$  ng/mL<sup>2</sup>) was selected based on its ability to identify clinically significant cancers missed on abbreviated MRI in ReIMAGINE [17, 18].

### Exploratory analyses

To assess the robustness of the ‘stage-gated’ approach, we performed sensitivity analyses exploring different combinations of readers at Stage 1 and criteria without the additional PSA<sub>d</sub> thresholds. These analyses aimed to evaluate whether modifying key pathway components would affect diagnostic accuracy. Full details are provided in the Supplementary Material.

### Outcome measures

The primary outcome was the PPV of the ‘stage-gated’ approach and bpMRI-adapted PI-RADS and Likert scores  $\geq 4$  from the original IP1-PROSTAGRAM MRI reports. Secondary outcomes included the number of recommended biopsies, detection of grade group (GG)  $\geq 2$  cancers, and benefit-to-harm ratios defined by Schoots et al [5]. Sensitivity, specificity, and negative predictive value are not reported due to the absence of reference standard verification in screen-negative participants.

The reference standard was IP1-PROSTAGRAM biopsy, with significant cancer defined as any length of GG  $\geq 2$  cancer (Gleason score  $\geq 3 + 4$ ) on systematic or targeted biopsy.

### Statistical analysis

No formal sample size calculation was performed as the study was designed as a secondary analysis.

Scans reported as non-diagnostic at Stage 1 were excluded from the analysis. Due to a high initial rate of non-diagnostic classifications, all non-diagnostic scans were re-read following reader retraining.

PPV was defined as the proportion of participants recommended for biopsy with GG  $\geq 2$  cancer. For the primary analysis, PPV was calculated in participants with reference standard verification (complete-case analysis) with Wilson 95% confidence intervals [23]. For the exploratory analyses, PPV was estimated using both complete-case analysis and methods to adjust for verification bias and missing reference standard data, including multiple imputation (standard and penalised logistic regression) as described by Day et al [24] (see Supplementary Material). Statistical analysis was performed by T.P. using Stata (v18.5, StataCorp).

## Results

### Characteristics

Of 408 consented IP1-PROSTAGRAM participants (median age 57 years [IQR 53, 61]), 405 had scans available for secondary analysis. Participant characteristics are shown in Table 2. The prevalence of GG  $\geq 2$  cancer was 4% (17/405).

### Stage-gated reporting outcomes

Following review of limited MRI sequences in Stage 1, 30 of 405 (7%) scans met criteria for Stage 2 evaluation - 23

scans were deemed screen-positive by consensus of at least two radiologists, and 7 scans were screen-positive by one radiologist only with raised PSAd. An additional 49 of 405 (12%) were screen-positive by one radiologist only but had normal PSAd, therefore were not advanced to Stage 2. At Stage 1, 39 of 405 (10%) scans were initially classified as non-diagnostic by at least one radiologist. Following reader retraining, this decreased to 20 of 405 (5%), which were excluded from further analysis (Supplementary Fig. S1).

**Table 2** Characteristics of IP1-PROSTAGRAM participants

Characteristic	Participants N = 408
Age (years)	57 (53, 61)
First-degree relative with PCa	
Present	43 (11)
Absent	360 (88)
Missing	5 (1)
PSA (ng/mL)	0.92 (0.57, 1.80)
PSA density (ng/mL <sup>2</sup> )	0.04 (0.02, 0.06)
Prostate volume (mL)	28 (22, 35)

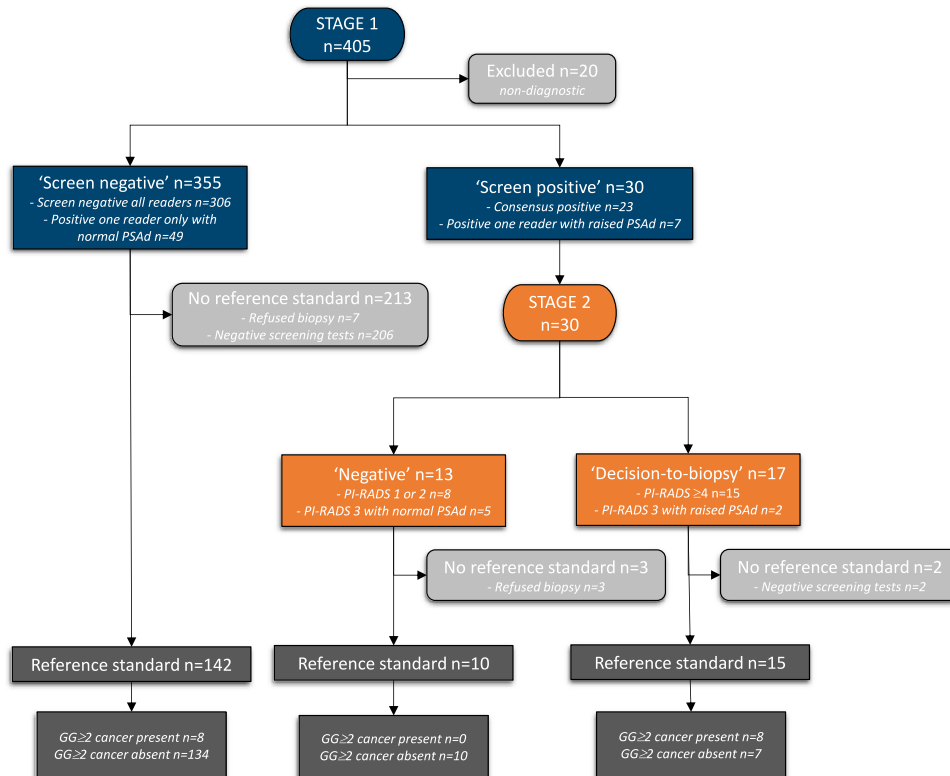
Data are n (%) or median (IQR)  
PSA prostate-specific antigen

The number of screen-positive scans at Stage 1 varied between readers (range 14–72; Table S3). Overall agreement between all three readers was 82% (95% CI: 78, 86; 316/385).

Following full bpMRI evaluation in Stage 2, 17 of 30 (57%) men had a hypothetical biopsy recommendation—15 based on PI-RADS score  $\geq 4$  and 2 based on PI-RADS score 3 with raised PSAd. Of these men, 8 of 17 (47%) had GG  $\geq 2$  cancer, 1 of 17 (6%) had GG 1 cancer, 6 of 17 (35%) had no cancer, and 2 of 17 (12%) had no available reference standard outcome. Figure 2 shows the screening flowchart. The GG  $\geq 2$  to GG 1 ratio, which reflects the ability of the screening strategy to detect clinically significant cancer (GG  $\geq 2$ ) while minimising the detection of insignificant cancers (GG 1), was 8:1. The ‘stage-gated’ reporting approach demonstrated improved benefit-to-harm ratios [5] compared to both Likert and PI-RADS  $\geq 4$  pathways based on the original IP1-PROSTAGRAM scores (Table 3).

**Diagnostic accuracy**

The PPV of the ‘stage-gated’ reporting pathway was 53% (95% CI: 30, 75; 8/15), compared to 29% (95% CI: 15, 47; 8/28) for the Likert  $\geq 4$ , and 30% (95% CI: 17, 46; 11/37)



**Fig. 2** Flowchart of the ‘stage-gated’ pathway. GG  $\geq 2$  grade group  $\geq 2$  cancer, i.e. clinically significant cancer, PSAd prostate-specific antigen density

**Table 3** Benefit-to-harm ratios [5] of the ‘stage-gated’ pathway compared to bp PI-RADS and Likert scores of  $\geq 4$  from the original IP1-PROSTAGRAM bpMRI reports

Outcome	Screening pathway (N = 385)		
	‘Stage-gated’ approach	bpPI-RADS $\geq 4$	Likert $\geq 4$
MRI-avoided biopsy	368	343	352
MRI-recommend biopsy	17	42	33
Biopsy outcome			
No biopsy	2 (12)	5 (12)	5 (15)
Benign	6 (35)	21 (50)	17 (52)
GG 1	1 (6)	5 (12)	3 (9)
GG $\geq 2$	8 (47)	11 (26)	8 (24)
Benefit-to-harm ratios <sup>5</sup>			
GG $\geq 2$ /GG 1	8.0	2.2	2.7
GG $\geq 2$ /GG 1 + benign biopsy	1.1	0.4	0.4
1 + benign biopsy			
Avoided biopsy/benign biopsy	61.3	16.3	20.7

The GG  $\geq 2$ /GG 1 ratio reflects the ability of the screening strategy to selectively detect clinically significant cancers (GG  $\geq 2$ ), while minimising the detection of insignificant cancers (GG 1). The GG  $\geq 2$ /GG 1 + benign biopsy ratio reflects the overall efficiency of the screening strategy in detecting clinically significant cancer (GG  $\geq 2$ ) relative to the number of insignificant (GG 1) or benign biopsies. The avoided biopsy/benign biopsy ratio reflects the balance of men avoiding biopsy (in those with negative MRI results) and benign biopsies (in those with positive MRI results). Data are n or n (%). GG grade group, GG  $\geq 2$  clinically significant cancer, GG 1 clinically insignificant cancer

for the PI-RADS  $\geq 4$  pathways. The ‘stage-gated’ approach recommended biopsy in approximately half as many men as the Likert  $\geq 4$  and PI-RADS  $\geq 4$  pathways (17 vs 33 and 42, respectively) (Fig. 3).

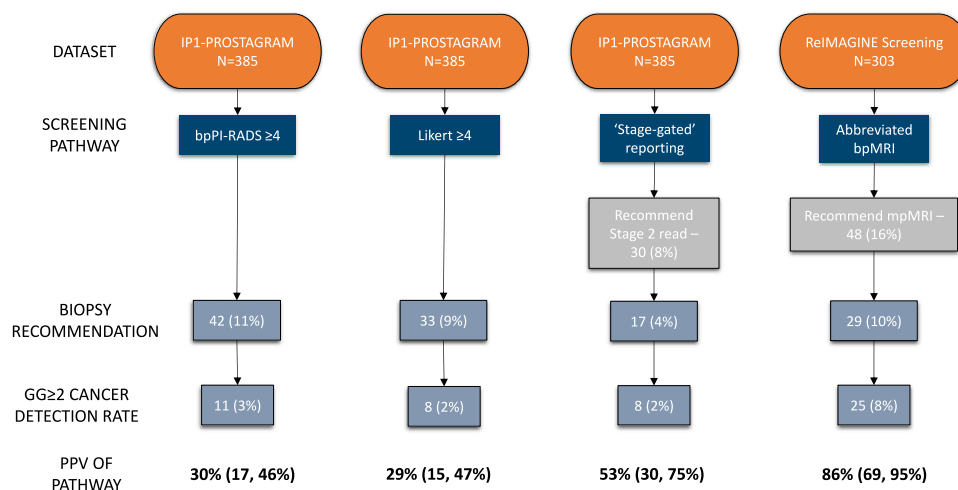
Of 17 men with GG  $\geq 2$  cancer within the entire cohort, the PI-RADS  $\geq 4$  pathway detected 11 cases, while the ‘stage-gated’ and Likert  $\geq 4$  pathways both detected 8. The three GG  $\geq 2$  cancers identified by PI-RADS  $\geq 4$  but missed by the ‘stage-gated’ pathway did not advance to Stage 2 (two were screen-positive by only one radiologist with normal PSA, and one was screen-negative by all radiologists) (Fig. 4). Of the remaining six men with GG  $\geq 2$  cancer, five were classified as screen-negative by all radiologists at Stage 1, and one was classified as non-diagnostic.

Among the men with PI-RADS or Likert scores  $\geq 4$  who would have correctly avoided biopsy by being classified as consensus screen-negative at Stage 1, in a considerable proportion (9/20, 45%), the elevated PI-RADS score was driven by appearances on the ADC map (Supplementary Table S4).

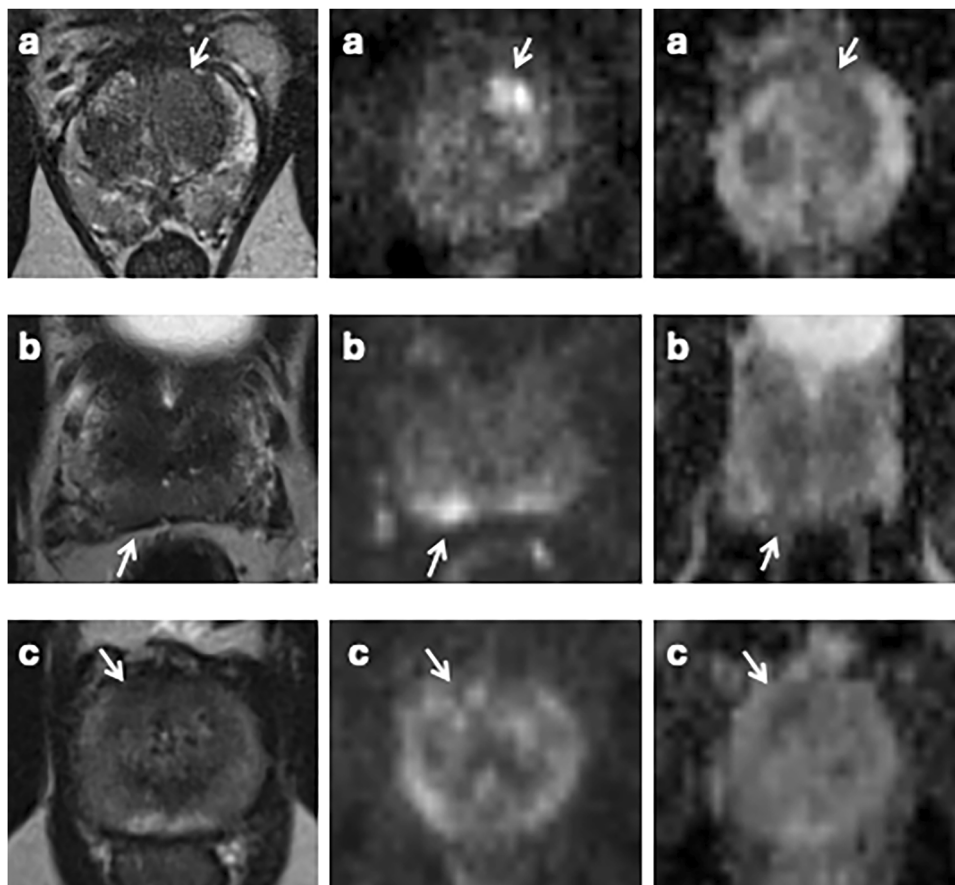
**Exploratory analyses**

Using multiple imputation to account for missing reference standard data in the two men who would have received a biopsy recommendation on ‘stage-gated’ reading but were negative in the original study, the PPV of the ‘stage-gated’ pathway was 47% (95% CI: 26, 69).

The PPVs of alternative screening pathways are detailed in the Supplementary Material (Tables S6–S9). PPV was



**Fig. 3** Diagnostic performance of the ‘stage-gated’ reporting approach compared to bpPI-RADS and Likert scores of  $\geq 4$  from the original IP1-PROSTAGRAM bpMRI reports. Diagnostic performance of the ReIMAGINE screening pathway [18] is included for comparison. PPV, positive predictive value; GG  $\geq 2$ , grade group  $\geq 2$  cancer, i.e. clinically significant cancer. Data are n (%) or % (95% CI)



**Fig. 4** Three cases (a–c) of GG  $\geq 2$  PCa detected by the bpPI-RADS  $\geq 4$  pathway but missed by both the ‘stage-gated’ reading and Likert  $\geq 4$  pathways. The GG  $\geq 2$  lesion (arrow) is shown on axial T2WI (left), high  $b$ -value ( $b1500$ ) DWI (middle), and ADC maps (right). Scans were all performed at 3-T. **a** A 63-year-old man with PSA 1.7 ng/mL and PSA density (PSAd) 0.05 ng/mL<sup>2</sup>. **b** A 58-year-old man with PSA 1.0 ng/mL and PSAd 0.03 ng/mL<sup>2</sup>. Both (a, b) were positive by only one radiologist in Stage 1. Due to normal PSAd, Stage 2 review was not performed. These cases were classified as radiologist interpretation errors. **c** A 59-year-old man with PSA 1.8 ng/mL and PSAd 0.05 ng/mL<sup>2</sup>. The screening images were negative by all radiologists in Stage 1, so they did not have Stage 2 review. In retrospect, the lesion is mainly conspicuous on the ADC map. This case was classified as a limitation of the reading strategy

comparable for pathways with two readers with a third acting as an arbitrator (PPV 55%, 6/11, 95% CI: 28, 79); however, the inclusion of the PSAd threshold at Stage 1 detected two additional GG  $\geq 2$  cancers. In contrast, pathways relying on single-reader decisions showed greater variability in PPV.

### Discussion

The ‘stage-gated’ approach for bpMRI reading demonstrated a higher PPV than adapted Likert and PI-RADS scoring, achieving a PPV of 53% (95% CI: 30, 75), compared to 29% (95% CI: 15, 47) and 30% (95% CI: 17, 46) for the Likert and PI-RADS  $\geq 4$  pathways, respectively. This approach halved the number of recommended biopsies while maintaining similar cancer detection rates and improving benefit-to-harm ratios.

Minimising harm is a fundamental requirement of screening programmes, making the reduction of false positives and unnecessary biopsies critical to optimising the balance of benefits and harms. In line with recent PCa screening literature [5, 7], we focused on PPV and benefit-to-harm ratios as key diagnostic accuracy measures, as these can be robustly estimated in a screening context with biopsy verification.

The ‘stage-gated’ approach was designed to replicate the ReIMAGINE Screening [17] pathway using a single bpMRI examination. The PPV of ‘stage-gated’ reporting (53%) was lower than that of the overall ReIMAGINE pathway (86%, 95% CI: 69, 95), which incorporated mpMRI in screen-positive cases, and slightly lower than the abbreviated bpMRI (59%, 95% CI: 44, 72) [18]. This may reflect the lower prevalence of GG  $\geq 2$  cancers in our

cohort (4% vs 10%), which directly impacts PPV, as well as the use of b1500 rather than b2000 DWI at 3-T. Nevertheless, the PPV of 'stage-gated' reading exceeded traditional scoring methods in IP1-PROSTAGRAM [12] and was higher than the PI-RADS  $\geq 4$  threshold reported in VISIONING [14] (28%, 95% CI: 19, 40) and MVP [13] screening studies (46%, 95% CI: 28, 65).

Among men with PI-RADS and Likert scores  $\geq 4$  who were correctly screened out at Stage 1, elevated scores were often driven by ADC findings with limited conspicuity on high *b*-value DWI, highlighting the importance of the high *b*-value DWI sequence [25]. In 'stage-gated' reading, a focal abnormality scoring  $\geq 4/5$  must be present on high *b*-value DWI to be considered screen-positive, which sets a higher threshold for positivity and reduces false positives. This may help to address the lower PPV of MRI in screening and improve the benefit-to-harm ratio by reducing unnecessary biopsies. A greater preponderance of GG  $\geq 2$  to GG 1 cancers may further reduce harm by improving histological selectivity.

The PI-RADS  $\geq 4$  pathway detected three additional GG  $\geq 2$  cancers compared to 'stage-gated' reporting. Two were likely due to radiologist interpretation error, and one was more conspicuous on the ADC map (a limitation of stage-gated reading). As this was a retrospective study, cases may have been more likely to be missed than in clinical practice, where reporting directly influences management. Other missed GG  $\geq 2$  cancers did not meet the MRI positivity threshold (no lesions scored  $\geq 4/5$ ). These missed cases are consistent with GÖTEBORG-2 [26] and STHLM3-MRI [27], where MRI served as a second-stage test following raised PSA. Repeated screening rounds may help mitigate the harms of missed cancers.

While incorporating MRI into the screening pathway may reduce the harms of PSA-based screening, MRI is costly and requires expert interpretation. 'Stage-gated' reading may improve reporting efficiency by allowing most scans to be assessed using only limited sequences, reducing reading time. Retrospective review of ReIMAGINE Screening and IP1-PROSTAGRAM datasets showed all serious incidental findings were visible on axial T2WI and high *b*-value DWI sequences, suggesting a low risk of harm from missed incidental pathology (Supplementary Table S10).

In our screening criteria, adapted from ReIMAGINE [18], scans with one non-diagnostic sequence were considered screen-negative if no suspicious lesion ( $\geq 4/5$ ) was seen on the other sequence. If a suspicious lesion was present, the scan was classified as non-diagnostic, and the participant would be recalled. This approach differs from conventional reporting and requires specific training, as reflected by the re-read in our study. The higher threshold

for a suspicious lesion on T2WI when b1500 DWI was non-diagnostic ( $\geq 4/5$  vs  $\geq 3/5$  for screen-positive) was intended to minimise unnecessary recalls by reserving repeat imaging for higher-suspicion cases. In ReIMAGINE, half of the false-positive scans had equivocal (3/5) T2WI changes compared with only 15% of true-positives [18], underscoring the need to balance recall reduction with cancer detection. Our recall rate (5%) was higher than ReIMAGINE (<1%) [18], though comparable screening data are lacking. In diagnostic settings, non-diagnostic (PI-QUAL <3) rates of 5–11% have been reported [28, 29]. The lower recall rate in ReIMAGINE may reflect its use of a single 3-T scanner, highlighting the need for robust quality control in screening MRI implementation. Finally, scans were classified as non-diagnostic if any of the three readers reported them as such, reflecting a conservative approach. Further evaluation is needed to determine whether advancing such cases to Stage 2, when all sequences are available, could reduce recall rates.

Despite standardised reporting criteria, inter-reader variability was observed at Stage 1. While all readers were expert urologists, tailored training for screening contexts—where disease prevalence differs from diagnostic populations—may improve inter-reader agreement. Although three readers reviewed scans concurrently, this configuration is functionally equivalent to a two-reader workflow with a third acting as an arbitrator, consistent with established double-reading models in population screening [30]. In clinical practice, a third reader would only be required in cases of disagreement; with 82% agreement between all three readers, this would be approximately 18% of cases. Two-reader pathways with a third as an arbitrator demonstrated comparable PPV, although the inclusion of the PSA<sub>d</sub> threshold at Stage 1 slightly improved GG  $\geq 2$  cancer detection. Further research could explore the role of artificial intelligence as a second reader, as is being investigated in breast cancer screening [31, 32].

This study has limitations. First, as a retrospective study, 'stage-gated' reporting did not influence biopsy decisions. While our findings demonstrate the feasibility of the approach and suggest a higher PPV over traditional scoring methods, prospective validation is needed. Data from the prospective ReIMAGINE study [18] support the potential success of such an approach. Second, screen-negative and some screen-positive men did not undergo disease status verification, introducing potential verification bias. Accordingly, we report biopsy-verified measures (e.g., PPV), which can be robustly estimated in a screening context. To address potential bias from missing reference standard data, complementary multiple imputation analyses were performed. This accounted for missing

outcomes, including two men recommended for biopsy on ‘stage-gated’ reading but negative in the original study, and yielded consistent PPV estimates. Lastly, the small number of  $GG \geq 2$  cancers (4%) in this screening cohort led to wide confidence intervals, limiting statistical precision and potentially reducing the generalisability of our findings despite observed PPV improvements. Consequently, the study was not powered to formally test non-inferiority, and comparisons should be interpreted descriptively.

In conclusion, the ‘stage-gated’ approach for reporting prostate screening MRI demonstrated higher PPV than conventional Likert/PI-RADS scoring, halving unnecessary biopsies while maintaining comparable cancer detection. These findings suggest that using limited MRI sequences for the initial read may improve the benefit-to-harm ratio of MRI-based screening, providing supportive data for possible implementations in cancer screening trials.

#### Abbreviations

ADC	Apparent diffusion coefficient
bpMRI	Biparametric MRI
DWI	Diffusion-weighted imaging
GG	Grade group
mpMRI	Multiparametric MRI
PCa	Prostate cancer
PPV	Positive predictive value
PSA	Prostate-specific antigen
PSAd	Prostate-specific antigen density
T2WI	T2-weighted imaging

#### Supplementary information

The online version contains supplementary material available at <https://doi.org/10.1007/s00330-025-12250-4>.

#### Acknowledgements

This study is a secondary analysis of data from the IP1-PROSTAGRAM study. The original study was sponsored by Imperial College London and funded by a Wellcome Trust Programme Grant (grant number 204998/Z/16/Z), a BMA Foundation for Medical Research grant, a Urology Foundation Research Scholarship (RESCH18), a Royal College of Surgeons of England Research Fellowship, and a grant from UK National Institute of Health Research Imperial Biomedical Research Centre. The funders had no role in the design and conduct of the IP1-PROSTAGRAM study.

#### Funding

S.P. is funded by a National Institute for Health and Care Research (NIHR) Research Professorship Award (NIHR303144).

#### Compliance with ethical standards

##### Guarantor

The scientific guarantor of this publication is Prof. Shonit Punwani.

##### Conflict of interest

The authors of this manuscript declare relationships with the following companies: F.G. reports consulting fees from SpectraCure and Procept outside of the submitted work. He has received speaker fees from Bayer and Siemens, is a member of the Scientific Editorial Board of *European Radiology* (section: Urological), and the Guest Editor of the *Active Surveillance of Prostate Cancer: MRI and beyond* Collection; as such, he has not participated in the selection or

review processes for this article. S.P. reports consulting fees from NVISION IMAGING, Gold Standard Phantoms and Quibim outside of the submitted work.

#### Statistics and biometry

One of the authors (T.P.) has significant statistical expertise.

#### Informed consent

Written informed consent was obtained from all subjects (patients) in the original IP1-PROSTAGRAM study. The original study consent form included a specific optional clause allowing the use of data for future retrospective research. The current study only includes participants who provided this consent.

#### Ethical approval

Institutional Review Board approval was obtained for the original IP1-PROSTAGRAM study. Ethical approval was granted by the UK National Research Ethics Committee (reference: 8/LO/1338).

#### Study subjects or cohorts overlap

Participants from IP1-PROSTAGRAM have been included in prior publications: Eldred-Evans et al [19], Eldred-Evans et al [12], and Day et al [24].

#### Methodology

- Retrospective
- Diagnostic study
- Multicentre

#### Author details

<sup>1</sup>Centre for Medical Imaging, University College London, London, UK. <sup>2</sup>Department of Radiology, University College London Hospital NHS Foundation Trust, London, UK. <sup>3</sup>Clinical and Experimental Radiology Unit, Experimental Imaging Center, IRCCS San Raffaele Scientific Institute, Milan, Italy. <sup>4</sup>Vita-Salute San Raffaele University, Milan, Italy. <sup>5</sup>Division of Surgical and Interventional Science, University College London, London, UK. <sup>6</sup>Department of Radiology, Addenbrooke’s Hospital and University of Cambridge, Cambridge, UK. <sup>7</sup>Division of Surgery, Imperial College London, London, UK. <sup>8</sup>Imperial Urology, Imperial College Healthcare NHS Trust, London, UK. <sup>9</sup>Department of Radiology, Imperial College Healthcare NHS Trust, London, UK. <sup>10</sup>Department of Radiology, The Hillingdon Hospitals NHS Foundation Trust, London, UK. <sup>11</sup>Paul Strickland Scanner Centre, Mount Vernon Hospital, Middlesex, UK. <sup>12</sup>Department of Urology, University College London Hospitals NHS Foundation Trust, London, UK.

Received: 25 August 2025 Revised: 3 November 2025 Accepted: 15 November 2025

Published online: 19 February 2026

#### References

1. Publications Office of the European Union (2022) Cancer screening in the European Union. <https://data.europa.eu/doi/10.2777/867180>. Accessed 1 Oct 2025
2. Hugosson J, Roobol MJ, Månsson M et al (2019) A 16-yr follow-up of the European randomized study of screening for prostate cancer. *Eur Urol* 76:43–51. <https://doi.org/10.1016/j.eururo.2019.02.009>
3. Thompson IM, Pauler DK, Goodman PJ et al (2004) Prevalence of prostate cancer among men with a prostate-specific antigen level  $\leq 4.0$  ng per milliliter. *N Engl J Med* 350:2239–2246. <https://doi.org/10.1056/nejmoa031918>
4. Martin RM, Donovan JL, Turner EL et al (2018) Effect of a low-intensity PSA-based screening intervention on prostate cancer mortality: the CAP randomized clinical trial. *JAMA* 319:883–895. <https://doi.org/10.1001/jama.2018.0154>
5. Schoots IG, Haider MA, Punwani S, Padhani AR (2025) MRI in prostate cancer screening: a review and recommendations, from the AJR special

- series on screening. *AJR Am J Roentgenol.* <https://doi.org/10.2214/AJR.24.32588>
6. Fazekas T, Shim SR, Basile G et al (2024) Magnetic resonance imaging in prostate cancer screening: a systematic review and meta-analysis. *JAMA Oncol* 10:745–754. <https://doi.org/10.1001/jamaoncol.2024.0734>
  7. Schoots IG, Ahmed HU, Albers P et al (2025) Magnetic resonance imaging-based biopsy strategies in prostate cancer screening: a systematic review. *Eur Urol* 88:247–260. <https://doi.org/10.1016/j.eururo.2025.05.038>
  8. van der Leest M, Israël B, Cornel EB et al (2019) High diagnostic performance of short magnetic resonance imaging protocols for prostate cancer detection in biopsy-naïve men: the next step in magnetic resonance imaging accessibility. *Eur Urol* 76:574–581. <https://doi.org/10.1016/j.eururo.2019.05.029>
  9. Alabousi M, Salameh JP, Gusenbauer K et al (2019) Biparametric vs multiparametric prostate magnetic resonance imaging for the detection of prostate cancer in treatment-naïve patients: a diagnostic test accuracy systematic review and meta-analysis. *BJU Int* 124:209–220. <https://doi.org/10.1111/bju.14759>
  10. Twilt JJ, Saha A, Bosma JS et al (2025) Evaluating biparametric versus multiparametric magnetic resonance imaging for diagnosing clinically significant prostate cancer: an international, paired, noninferiority, confirmatory observer study. *Eur Urol* 87:240–250. <https://doi.org/10.1016/j.eururo.2024.09.035>
  11. Ng ABCD, Asif A, Agarwal R et al (2025) Biparametric vs multiparametric MRI for prostate cancer diagnosis: the PRIME diagnostic clinical trial. *JAMA* 334:1170–1179. <https://doi.org/10.1001/jama.2025.13722>
  12. Eldred-Evans D, Tam H, Sokhi H et al (2023) An evaluation of screening pathways using a combination of magnetic resonance imaging and prostate-specific antigen: results from the IP1-PROSTAGRAM study. *Eur Urol Oncol* 6:295–302. <https://doi.org/10.1016/j.euro.2023.03.009>
  13. Nam R, Patel C, Milot L et al (2022) Prostate MRI versus PSA screening for prostate cancer detection (the MVP Study): a randomised clinical trial. *BMJ Open* 12:e059482. <https://doi.org/10.1136/bmjopen-2021-059482>
  14. Wetterauer C, Matthias M, Püeschel H et al (2024) Opportunistic prostate cancer screening with biparametric magnetic resonance imaging (VISIONING). *Eur Urol Focus* 10:332–338. <https://doi.org/10.1016/j.euf.2024.02.006>
  15. Messina E, La Torre G, Pecoraro M et al (2024) Design of a magnetic resonance imaging-based screening program for early diagnosis of prostate cancer: preliminary results of a randomized controlled trial—prostate cancer secondary screening in Sapienza (PROSA). *Eur Radiol* 34:204–213. <https://doi.org/10.1007/s00330-023-10019-1>
  16. Marsden T, Lomas DJ, McCartan N et al (2021) ReIMAGINE prostate cancer screening study: protocol for a single-centre feasibility study inviting men for prostate cancer screening using MRI. *BMJ Open* 11:e048144. <https://doi.org/10.1136/bmjopen-2020-048144>
  17. Moore CM, Frangou E, McCartan N et al (2023) Prevalence of MRI lesions in men responding to a GP-led invitation for a prostate health check: a prospective cohort study. *BMJ Oncol* 2:e000057. <https://doi.org/10.1136/bmjonc-2023-000057>
  18. Thorley N, Parry T, Giganti F et al (2025) Diagnostic accuracy of abbreviated biparametric MRI for prostate cancer screening: a prospective feasibility study (ReIMAGINE Study). *Eur Radiol.* <https://doi.org/10.1007/s00330-025-11837-1>
  19. Eldred-Evans D, Burak P, Connor MJ et al (2021) Population-based prostate cancer screening with magnetic resonance imaging or ultrasonography: the IP1-PROSTAGRAM study. *JAMA Oncol* 7:395–402. <https://doi.org/10.1001/jamaoncol.2020.7456>
  20. Weinreb JC, Barentsz JO, Choyke PL et al (2016) PI-RADS prostate imaging—reporting and data system: 2015, version 2. *Eur Urol* 69:16–40. <https://doi.org/10.1016/j.eururo.2015.08.052>
  21. Xie SW, Wang YQ, Dong BJ et al (2018) A nomogram based on a TRUS five-grade scoring system for the prediction of prostate cancer and high grade prostate cancer at initial TRUS-guided biopsy. *J Cancer* 9:4382–4390. <https://doi.org/10.7150/jca.27344>
  22. Turkbey B, Rosenkrantz AB, Haider MA et al (2019) Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2. *Eur Urol* 76:340–351. <https://doi.org/10.1016/j.eururo.2019.02.033>
  23. Wilson EB (1927) Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc* 22:209–212. <https://doi.org/10.1080/01621459.1927.10502953>
  24. Day E, Eldred-Evans D, Prevost AT et al (2022) Adjusting for verification bias in diagnostic accuracy measures when comparing multiple screening tests—an application to the IP1-PROSTAGRAM study. *BMC Med Res Methodol.* <https://doi.org/10.1186/s12874-021-01481-w>
  25. Godley KC, Syer TJ, Toms AP et al (2018) Accuracy of high *b*-value diffusion-weighted MRI for prostate cancer detection: a meta-analysis. *Acta Radiol* 59:105–113. <https://doi.org/10.1177/0284185117702181>
  26. Hugosson J, Godtman RA, Wallstrom J et al (2024) Results after four years of screening for prostate cancer with PSA and MRI. *N Engl J Med* 391:1083–1095. <https://doi.org/10.1056/NEJMoa2406050>
  27. Nordström T, Annerstedt M, Glaessgen A et al (2024) Repeated prostate cancer screening using prostate-specific antigen testing and magnetic resonance imaging: a secondary analysis of the STHLM3-MRI randomized clinical trial. *JAMA Netw Open* 7:e2354577. <https://doi.org/10.1001/jamanetworkopen.2023.54577>
  28. Giganti F, Allen C, Emberston M et al (2020) Prostate imaging quality (PI-QUAL): a new quality control scoring system for multiparametric magnetic resonance imaging of the prostate from the PRECISION trial. *Eur Urol Oncol* 3:615–619. <https://doi.org/10.1016/j.euro.2020.06.007>
  29. Windisch O, Benamran D, Dariane C et al (2023) Role of the prostate imaging quality PI-QUAL score for prostate magnetic resonance image quality in pathological upstaging after radical prostatectomy: a multi-centre European study. *Eur Urol Open Sci* 47:94–101. <https://doi.org/10.1016/j.euro.2022.11.013>
  30. Taylor P, Potts HWW (2008) Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate. *Eur J Cancer* 44:798–807. <https://doi.org/10.1016/j.ejca.2008.02.016>
  31. Lång K, Josefsson V, Larsson AM et al (2023) Artificial intelligence-supported screen reading versus standard double reading in the mammography screening with artificial intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. *Lancet Oncol* 24:936–944. [https://doi.org/10.1016/S1470-2045\(23\)00298-X](https://doi.org/10.1016/S1470-2045(23)00298-X)
  32. NIHR National Institute for Health and Care Research (2025) World-leading AI trial to tackle breast cancer launched. <https://www.nihr.ac.uk/news/world-leading-ai-trial-tackle-breast-cancer-launched>. Accessed 7 Mar 2025

## Publisher's Note

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