

REVIEW

Magnetic resonance imaging on disease reclassification among active surveillance candidates with low-risk prostate cancer: a diagnostic meta-analysis

R Guo¹, L Cai¹, Y Fan, J Jin, L Zhou and K Zhang

BACKGROUND: Active surveillance (AS) is an increasingly important attempt to avoid overtreatment of patients who harbor clinically insignificant disease while offering curative treatment to those in whom disease is reclassified as higher risk after an observation period and repeat biopsy. We aim to evaluate the diagnostic performance of magnetic resonance imaging (MRI) in predicting upgrading on confirmatory biopsy in men with low-risk prostate cancer (PCa) on AS.

METHODS: We searched the PubMed for pertinent studies up to November 2014. We used standard methods recommended for meta-analyses of diagnostic test evaluations. The analysis was based on a summary receiver operating characteristic (SROC) curve. Meta-regression analysis was used to assess the effects of some confounding factors on the results of the meta-analysis. The potential presence of publication bias was tested using the Deeks' funnel plots.

RESULTS: Seven studies provided the diagnostic data on MRI and AS of PCa, comprising 1028 patients. The pooled estimates of MRI on disease reclassification among AS candidates were as follows: sensitivity, 0.69 (95% confidence interval (CI), 0.44–0.86); specificity, 0.78 (95% CI, 0.53–0.91); positive likelihood ratio, 3.1 (95% CI, 1.6–6.0); negative likelihood ratio, 0.40 (95% CI, 0.23–0.70); and diagnostic odds ratio, 8 (95% CI, 4–16). The *P*-value for heterogeneity was < 0.001 . We found that the SROC curve is positioned toward the desirable upper left corner of the curve, and the area under the curve was 0.79 (95% CI, 0.76–0.83). For a pretest probability of 0.20, the corresponding positive predictive value (PPV) was 0.44 and the negative predictive value (NPV) was 0.91. MRI may reveal an unrecognized significant lesion in 33.27% of patients, and biopsy of these areas reclassified 14.59% of cases as no longer fulfilling the criteria for AS. In addition, when no suspicious disease progression (66.34%) was identified on MRI, the chance of reclassification on repeat biopsy was extremely low at 6.13%.

CONCLUSIONS: MRI, especially multiparametric (MP)-MRI, has a moderate diagnostic accuracy as a significant predictor of disease reclassification among AS candidates. The high NPV and specificity for the prediction of biopsy reclassification upon clinical follow-up suggest that negative prostate MRI findings may support a patient remaining under AS. Although the PPV and sensitivity for the prediction were relatively low, the presence of a suspicious lesion > 10 mm lesion may suggest an increased risk for disease progression.

Prostate Cancer and Prostatic Disease (2015) **18**, 221–228; doi:10.1038/pcan.2015.20; published online 19 May 2015

INTRODUCTION

The incidence and prevalence of prostate cancer (PCa) have both increased significantly over the past two decades, and PCa has surpassed lung cancer as the most common cancer in men. It is generally accepted that these changes resulted from PSA screening. At the same time, PCa-specific mortality has decreased, which is related to early detection and treatment. Nevertheless, treatment of low-risk PCa results in unnecessary side effects, which impair the quality of life of patients and their families, and healthcare expenses.^{1–2} Active surveillance (AS) is an increasingly important attempt to avoid overtreatment of patients who harbor clinically insignificant disease while offering curative treatment to those in whom disease is reclassified as higher risk after an observation period and repeat biopsy.^{3–6}

Up to 35% of patients in AS will experience biopsy reclassification at a median follow-up of 2.9 years.^{7–9} Unfortunately,

noninvasive methods of cancer surveillance, including PSA doubling time (area under the curve (AUC) 0.59) and PSA velocity (AUC 0.61), are not associated with prostate biopsy reclassification, and annual prostate biopsy is currently recommended by some for monitoring men undergoing AS.⁹ Nevertheless, serial prostate biopsies are associated with potentially serious infectious and quality-of-life sequelae.^{10–11} Improved patient selection and less invasive methods of cancer surveillance are needed to improve the safety of AS in the management of low-risk PCa.

Magnetic resonance imaging (MRI) can provide additional high-resolution information on tissue properties, such as diffusion and enhancement. MRI enables soft tissue contrast and characterization, and advanced computational methods are available to assess function. Thus, MRI alone or combined with clinical parameters may be useful in the prediction of insignificant PCa, particularly in the context of clinically nonpalpable tumors.^{12–14} Previously,

Department of Urology, Peking University First Hospital and Institute of Urology, National Research Center for Genitourinary Oncology, Beijing, China. Correspondence: Dr. K Zhang, Department of Urology, Peking University First Hospital and Institute of Urology, National Urological Cancer Center, 8 Xishiku Street, Xicheng District, Beijing 100034, China.

E-mail: kaizhang.pku@gmail.com

¹These authors contributed equally to this work.

Received 20 December 2014; revised 28 March 2015; accepted 31 March 2015; published online 19 May 2015

studies have shown that MRI may be a promising imaging technique in identifying men entering AS and monitoring these men in the AS cohort. This study is a diagnostic meta-analysis that aims to summarize the diagnostic performance of MRI on PCa reclassification among AS candidates.

MATERIALS AND METHODS

Selection of studies

All authors participated in the selection of eligible studies for inclusion. We reviewed PubMed for citations published before November 2014, describing MRI used in AS for PCa. The search strategy included the terms prostate neoplasms (MeSH) or PCa, watchful waiting (MeSH) or AS or expectant management, reclassification or histologic progression, and MRI. Article titles and abstracts were reviewed independently for eligibility by two authors, and discrepancies were resolved by consensus. Studies were included if they met all the following criteria: (1) the study population consisted of PCa patients for AS; (2) the study assessed diagnostic performance of MRI for histologic progression; (3) cc analysis was used as the reference standard test; and (4) if at least one pair of the absolute numbers of true-positive results and false-negative results or true-negative results and false-positive results were available or could be derived adequately. To include true-positive results, false-positive results, true-negative results and false-negative results in a meta-analysis, all four should be available. Studies were excluded if they met one of the following criteria: (1) the article was a review or meta-analysis and (2) (potentially) overlapping study populations were reported for the same outcome.

Data extraction

Two reviewers independently screened all titles identified in the database searches. The full text of all articles included by either reviewer on the basis of the abstract was obtained. To determine eligibility for inclusion, one author reviewed all full-text articles. A second author repeated this assessment independently for a random selection of 10% of full-text articles, and there was complete agreement regarding the excluded articles. Two reviewers independently extracted data from all of the included studies. One author extracted all data from all studies, and the other author independently re-extracted data from all of the included studies between them. Discrepancies were resolved by consensus. Full-text copies of potentially relevant studies were obtained and their eligibility for inclusion was independently assessed. Studies that do not fulfill all of the inclusion criteria were excluded.

Quality assessment

We used the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) checklist to assess the study quality in terms of the risk of bias and the applicability of included studies.^{15–16} Each study was independently assessed by two reviewers. Any discrepancies were resolved by discussion or, if agreement could not be reached, by arbitration by a third reviewer.

Data analysis and synthesis

We used standard methods recommended for meta-analyses of diagnostic test evaluations. The following measures of test accuracy were computed for each study: sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR), and their 95% confidence interval (CI). The advantage of DOR is its independence from disease prevalence and the approximately normal distribution of the natural logarithm of DOR.¹⁷ As a general rule, diagnostic tests with a DOR > 25 are considered moderately accurate, and tests with a DOR > 100 are considered highly accurate.^{18–19}

If the 2×2 data were not available, attempts were made to derive them from reported summary statistics, such as the sensitivity, specificity and/or likelihood ratios.

Preliminary exploratory analyses were conducted for each test by plotting estimates of the sensitivity and specificity from each study on forest plots and in the receiver operating characteristic (ROC) space. The AUC, the corresponding 95% CI and the *P*-value were calculated.

The sensitivity and specificity for the single test threshold was used to plot a summary receiver operating characteristic (SROC) curve.²⁰

We adopted the following overall approach for the evaluation of heterogeneity in the results expected between studies of diagnostic tests. We assessed the heterogeneity using forest plots and then statistically tested for significance using the Chi-squared test and the *I*² statistic (*I*² > 50% indicates significant heterogeneity).²¹ Consequently, we calculated pooled estimates using bivariate mixed-effects binary regression modeling, which provides more conservative estimates than fixed-effects modeling when heterogeneity was present.²² Spearman's correlation analysis was used to check the threshold effect, which is conducive to heterogeneity. In addition, meta-regression was analyzed to identify possible sources of heterogeneity. Subgroup analyses and sensitivity analysis were also performed if necessary.

Publication bias is considered as a concern for meta-analyses of diagnostic studies. In our study, publication bias was assessed by the regression of the natural logarithm of DOR against the square root of the effective sample size; *P* < 0.05 for the slope coefficient is suggestive of significant publication bias.²³ The MIDAS module of STATA commands was used for the meta-analysis of the diagnostic data.^{24,25} STATA version 12 (StataCorp, College Station, TX, USA) and Meta-DiSc software (version 1.4; XI Cochrane Colloquium, Barcelona, Spain) were used for data analysis. Significance level was set to 0.05.

RESULTS

A total of 47 articles from PubMed were initially retrieved. The last update was on November 2014 (Figure 1). After reading the title and the abstract, 12 articles were found to be suitable for further evaluation. Of the 12 articles reviewed in detail, 5 were excluded (2 were reviews, 2 presented nonextractable data and 1 focused on expanding AS criteria), leaving 7 studies to be included in the final analysis (Table 1).^{26–32}

Study characteristics and quality assessment

After reading the full text, seven selected articles included 1028 patients. The samples were ranged from 50 to 388 patients. All of the seven included essays reported the mean age, which ranged

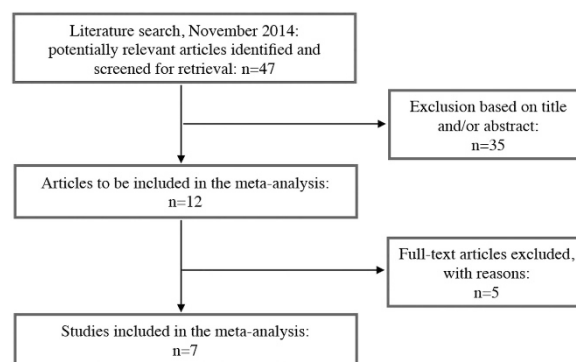


Figure 1. Flowchart of study selection.

Table 1. Summary of included studies and diagnostic estimates (individual study data and independent pooled sensitivity/specificity estimates)

Author	Country	Year	Mean age	AS protocols	TP	FP	FN	TN	AUC	Sensitivity, 95% CI	Specificity, 95% CI
Fradet <i>et al.</i> ²⁷	USA	2010	63.4	Low- and intermediate-risk clinical stage lower than T2c, Gleason score 6, PSA < 20 ng ml ⁻¹ and < 33% of biopsy cores positive for cancer	25	22	2	16	NA	0.93 (0.76–0.99)	0.42 (0.26–0.59)
Margel <i>et al.</i> ³⁰	Canada	2012	63.5	Low-risk T1c–T2a, Gleason score ≤ 6, PSA < 10 ng ml ⁻¹ , 2 ≤ positive cores, ≤ 50% cancer in each core	10	2	8	36	NA	0.56 (0.31–0.78)	0.95 (0.82–0.99)
Vargas <i>et al.</i> ²⁶	USA	2012	60.6	Low-risk T1–T2a, Gleason score ≤ 6, PSA < 10 ng ml ⁻¹	70	93	9	216	0.76–0.79	0.89 (0.79–0.95)	0.70 (0.64–0.75)
Stamatakis <i>et al.</i> ²⁸	USA	2013	60.2	Very low-risk T1c, Gleason score ≤ 6, PSA < 10 ng ml ⁻¹ , 2 ≤ positive cores, ≤ 50% cancer in each core, PSA density < 0.15 ng ml ⁻¹ g ⁻¹	22	36	3	24	0.71	0.88 (0.69–0.97)	0.40 (0.28–0.53)
Vasarainen <i>et al.</i> ²⁹	Finland	2013	64	Low-risk T stage 2 or below, Gleason score ≤ 6, 2 ≤ positive cores, PSA density < 0.2 ng ml ⁻¹ ml ⁻¹	12	28	11	29	NA	0.52 (0.31–0.73)	0.51 (0.37–0.64)
Mullins <i>et al.</i> ³¹	USA	2013	67	Very low-risk T1c, Gleason score ≤ 6, PSA < 10 ng ml ⁻¹ , 2 ≤ positive cores, ≤ 50% cancer in each core, PSA density < 0.15 ng ml ⁻¹ g ⁻¹	6	7	25	262	NA	0.19 (0.07–0.37)	0.97 (0.95–0.99)
Bonekamp <i>et al.</i> ³²	USA	2013	67	Very low-risk T1c, Gleason score ≤ 6, PSA < 10 ng ml ⁻¹ , 2 ≤ positive cores, ≤ 50% cancer in each core, PSA density < 0.15 ng ml ⁻¹ g ⁻¹	5	4	5	36	NA	0.50 (0.19–0.81)	0.90 (0.76–0.97)

Abbreviations: AS, active surveillance; AUC, area under the curve; CI, confidence interval; FP, false positive; FN, false negative; NA, not applicable; TP, true positive; TN, true negative.

from 60.2 to 67 years. Three of the seven studies included very low-risk PCa patients satisfying Johns Hopkins AS criteria,^{28,31,32} three included low-risk PCa patients^{26,29,30} and only Fradet *et al.*²⁷ identified patients with low- and intermediate-risk PCa. However, the AUC values were only reported in Vargas and Stamatakis's reports.^{26,28} The values of the MRI on disease reclassification and the basic characteristics (true-positive, false-positive, true-negative and false-negative values for MRI) were shown in Table 1, and the summary of included studies (MRI techniques, confirmatory biopsy and definition disease progression) was shown in Table 2 and Table 3.

The quality varied across included studies (summary of QUADAS-2 quality data, Figure 2 and Figure 3); however, all had low to unclear risk of bias and applicability concerns, and therefore we did not exclude any from analysis. None of the included studies had a case-control design. All studies had the same reference standard because histopathologic analysis was the inclusion criteria of this meta-analysis. The MRI interpretation was blinded in five studies, but this was unclear in the other two studies. Histopathologic interpretation was blinded in four studies, whereas this was unclear in three studies. We scored four of seven studies as having an appropriate interval between the MRI and biopsy, and we scored the other three studies as unclear because they did not report an interval at all.

Diagnostic accuracy

The sensitivity point estimates ranged from 0.19 to 0.93 across individual studies, and the specificity point estimates ranged from 0.40 to 0.97 (Figure 4). The pooled estimates of MRI for disease reclassification diagnosis were as follows: sensitivity, 0.69 (95% CI, 0.44–0.86); and specificity, 0.78 (95% CI, 0.53–0.91). Positive predictive value (PPV) (0.44) was relatively low, whereas negative predictive value (NPV) (0.91) was high, given a pretest probability of 20%. We also noted that the PLR was 3.1 (95% CI, 1.6–6.0), the NLR was 0.40 (95% CI, 0.23–0.70) and the DOR was 8 (95% CI, 4–16).

The SROC curve summarizes the overall diagnostic accuracy, showing the tradeoff between sensitivity and specificity. We found that the AUC was 0.79 (95% CI, 0.76–0.83) (Figure 5).

On the basis of the ROC plots, we identified one study as a potential outlier: the study by Hanna *et al.*²⁹ After exclusion of this potential outlier, the diagnostic accuracy of MRI for disease reclassification among AS candidates showed a sensitivity of 0.74 (95% CI, 0.45–0.91), a specificity of 0.81 (95% CI, 0.56–0.94) and a DOR of 13 (95% CI, 8–20).

Threshold effect and subgroup analysis

As different sensitivity and specificity by various research conditions cause different threshold effect³³ and DOR, it is necessary to assess the presence of a threshold effect. In addition, if a threshold effect exists, sensitivity and specificity show a negative correlation and the worker ROCplane scatter distribution is in a typical 'shoulder arm-shaped' style. In our study, the MRI of SROC curve does not show this typical style (Figure 5). It indicates that there is no heterogeneity from threshold effect, while the Spearman correlation coefficient was 0.679 and *P*-value was 0.094.

After testing for heterogeneity caused by other sources, the results showed that sensitivity (*P* < 0.001, *I*² = 91.3%), specificity (*P* < 0.001, *I*² = 96.9%), PLR (Cochrane *Q* = 59.04, *P* < 0.001, *I*² = 83.8%), NLR (Cochrane *Q* = 99.93, *P* < 0.001, *I*² = 94.00%) and DOR (Cochrane *Q* = 22.37, *P* = 0.001, *I*² = 73.2%) in the included studies showed high heterogeneity.

We consider the quality of the studies, AS protocols, year, country, MRI technique and biopsy strategy for the sources of the heterogeneity. Through meta-regression, we found that the *P*-value of the MRI with endorectal coils (ERCs) was 0.0182 and obviously less than others. The results suggested that the MRI technique might contribute to heterogeneity (Table 4).

Table 2. Summary of included studies (MRI techniques)

Author	Year	When MRI was performed	Type of MRI	MRI scoring system	MRI criterion
Fradet <i>et al.</i> ²⁷	2010	AS entry, 8-week wait time after biopsy	1.5T MRI with endorectal coil (T2W MRI and MRSI)	NA	Suspicious for cancer
Margel <i>et al.</i> ³⁰	2012	Within 1 year of diagnosis before repeat biopsy	1.5T MP-MRI with endorectal coil (T2W MRI, ADC maps of DW-MRI and DCE-MRI sequences)	NA	Suspicious lesions ≥ 10 mm in maximal diameter
Vargas <i>et al.</i> ²⁶	2012	Before biopsy	1.5T & 3T T2W-MRI with endorectal coil	5-point Likert scale	Score three or more
Stamatakis <i>et al.</i> ²⁸	2013	Before biopsy	3T MP-MRI with endorectal coil (T2W MRI, ADC maps of DW-MRI and DCE-MRI sequences)	National Cancer Institute lesion-based MP-MRI cancer suspicion scoring system: negative, low, moderate, high	Suspicion high level
Vasarainen <i>et al.</i> ²⁹	2013	After 1 year of surveillance before repeat biopsy	3T DW-MRI (ADC maps) with pelvic coil	NA	Anatomical lesion
Mullins <i>et al.</i> ³¹	2013	≥ 6 weeks after last biopsy	3T MP-MRI with endorectal coil (T2W MRI, ADC maps of DW-MRI and DCE-MRI sequences)	NA	Suspicious lesions ≥ 10 mm in maximal diameter
Bonekamp <i>et al.</i> ³²	2013	Before biopsy, or 8-week wait time after biopsy	3T MP-MRI with endorectal coil (T2W MRI, ADC maps of DW-MRI, DCE-MRI, and MRSI sequences)	NA	Suspicious lesions ≥ 10 mm in maximal diameter

Abbreviations: AS, active surveillance; ADC, apparent diffusion coefficient; DCE, dynamic contrast enhanced; DW, diffusion-weighted; MP, multiparametric; MRI, magnetic resonance imaging; MRSI, magnetic resonance spectroscopic imaging; NA, not applicable; T2W, T2-weighted.

A subgroup analysis was carried out using the following criteria—(1) AS candidates were divided into two groups according to AS protocols: three studies included very low-risk PCa patients and four studies included low-risk PCa; (2) studies included were divided into two groups according to year: three studies before 2012 or in 2012 and four studies after 2012; (3) patients were divided into two groups according to country: five studies from the USA and two studies outside the USA; (4) studies included were divided into two groups according to MRI technique: four in multiparametric (MP)-MRI group and three in non MP-MRI group, and six in the group with ERCs and one in the group without ERCs; (5) studies included were divided into two groups according to biopsy strategy: four studies with targeted biopsy and three studies without targeted biopsy. Only in studies with targeted biopsy and those without targeted biopsy there was significant difference in the sensitivity (0.80 vs 0.31, $P < 0.01$) and specificity (0.63 vs 0.95, $P = 0.02$).

Publication bias

The Deeks' funnel plots for publication bias²³ also showed no asymmetry (Figure 6). The evaluation of publication bias showed that the bias coefficient was -11.66 and that it was not significant ($P = 0.30$). There was no evidence that publication bias existed. However, because of the limited number of the articles, concluding whether the publication bias existed in this meta-analysis is difficult.

Clinical utility of index test

From the Fagan's nomogram, we found that when 20% was selected as the pretest probability the post-test probability would raise to 43% with a PLR of 3.1, and the probability would decrease to 9%, whereas the NLR was 0.40.

DISCUSSION

We have performed a meta-analysis of diagnostic accuracy of MRI on disease reclassification among AS candidates. Although meta-analysis is not yet a widely approved method to summarize evidence from diagnostic studies, we believe that pooling the diagnostic accuracy from eligible studies provides valuable information for urologists and researchers until larger studies are available.

We presented the PLR, NLR and DOR as our measures of diagnostic accuracy. $PLR > 10$ or $NLR < 0.1$ indicated high accuracy. $DOR > 25$ was considered moderately accurate, and $DOR > 100$ was considered highly accurate. A PLR value of 3.1 suggests that AS candidates with MRI suspicious disease progression have an approximately 3.1-fold higher chance of being MRI negative. In addition, the NLR was found to be 0.4; therefore, if the MRI result was negative, the probability that this AS candidate has disease progression that needs reclassification was $\sim 40\%$. A DOR of 8, although > 1 , indicates relatively poor performance in PCa reclassification.

Our relatively low PLR (3.1) and high NLR (0.4), combined with poor sensitivity (0.69) and specificity (0.78), do not give a sustained evidence of diagnostic accuracy of MRI on disease reclassification. These results suggest that the accuracy of MRI in the PCa reclassification may not be as high as previously described in some studies. In the study by Vargas *et al.*,²⁶ the sensitivity and specificity were 0.89 and 0.70, respectively. However, AUC value in our SROC curve was 0.79, indicating a moderate accuracy of the value of MRI in disease reclassification among AS candidates, as it is > 0.7 .³⁴ The relatively high specificity and AUC value demonstrated that a very low-risk or low-risk PCa with the negative MRI result supports no disease reclassification and continued AS.

Important statistical heterogeneity ($I^2 > 50\%$) was found in the analysis, which might be a confounding factor for the results. In

Table 3. Summary of included studies (confirmatory biopsy and disease progression)

Author	Year	Confirmatory biopsy	Disease progression
Fradet <i>et al.</i> ²⁷	2010	At least 14-core TRUS-guided prostate biopsy and targeted TRUS-guided biopsies	Gleason score >6, PSA velocity >0.75 (mg L ⁻¹) year ⁻¹ or initiation of treatment > 6 months after diagnosis
Margel <i>et al.</i> ³⁰	2012	12-core TRUS-guided systematic extended prostate biopsy and targeted TRUS-guided biopsies	Gleason score >6, > 2 positive cores, > 50% tumor in any core
Vargas <i>et al.</i> ²⁶	2012	Standard 12-core TRUS-guided biopsy, two biopsy samples obtained from the transition zone for a total of 14 cores, targeted biopsy at discretion	Gleason score >6
Stamatakis <i>et al.</i> ²⁸	2013	12-core TRUS-guided systematic extended prostate biopsy and targeted MRI-US fusion-guided biopsies	Gleason score >6, > 2 positive cores, > 50% tumor in any core
Vasarainen <i>et al.</i> ²⁹	2013	12-core TRUS-guided systematic extended prostate biopsy, targeted biopsy at discretion	Gleason score >6, > 2 positive cores, T stage > 2, PSA-DT < 3 years
Mullins <i>et al.</i> ³¹	2013	12-core systematic biopsies with or without two additional cores taken from the transition zone	Any individual biopsy Gleason pattern > 3 or sum > 6, > 2 positive cores, > 50% tumor in any core
Bonekamp <i>et al.</i> ³²	2013	12- to 14-core systemic biopsy including transition zone biopsies	Gleason score >6, > 2 positive cores, > 50% tumor in any core

Abbreviations: MRI-US, magnetic resonance imaging-ultrasound; PSA-DT, PSA doubling time; TRUS, transrectal ultrasound.

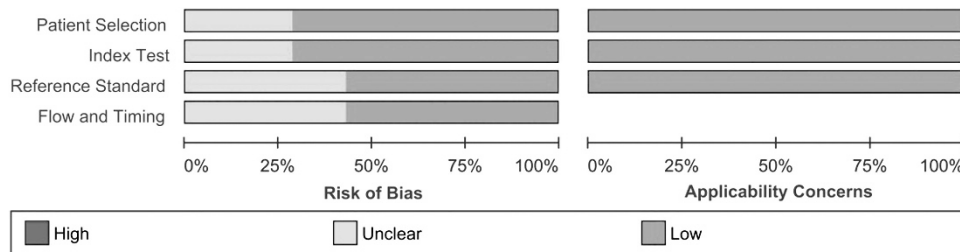


Figure 2. Graph showing the risk of bias and applicability concerns: review of authors' judgements about each domain, presented as percentages across included studies.

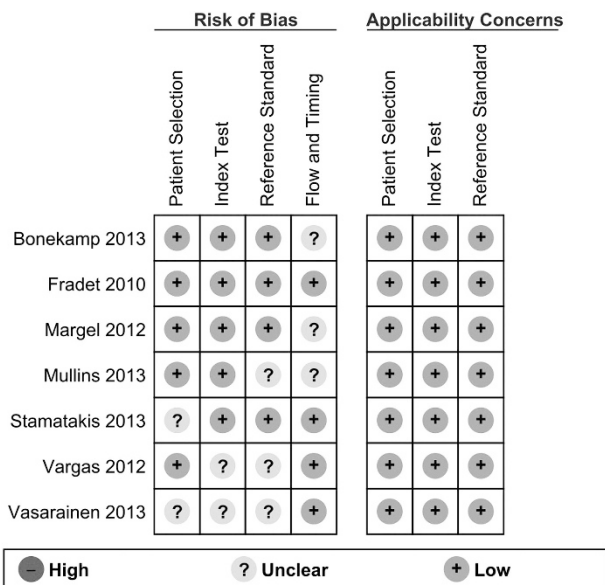


Figure 3. Chart summarizing the risk of bias and applicability concerns: review of authors' judgements about each domain for each included study. -, high concern; ?, unclear concern; +, low concern.

addition, the threshold effect must be considered first in test accuracy studies, which arises when differences exist in sensitivity and specificity owing to the threshold used in different studies to define a positive or negative test result.^{33,35} We use Spearman

correlation coefficient to analyze the threshold effect, and its value is 0.679 ($P=0.094$), which indicates that there is no heterogeneity from threshold effects. To further evaluate the sources of heterogeneity, meta-regression analysis is conducted in terms of quality of the studies, AS protocols, year, country, MRI technique and biopsy strategy. This analysis indicates that the results may be influenced by Hanna's study.²⁹ However, after this study is removed, the data analysis results do not change obviously.

In the subgroup analysis, only after dividing the seven included studies into two categories according to biopsy strategy, the sensitivity in the targeted biopsy category is significantly better than non-targeted biopsy category, with the exception of specificity, indicating that the value of MRI on PCa reclassification among AS candidates may be different whether it is targeted biopsy or not. Theoretically targeted biopsies from lesions based on clinical or imaging findings could have improved tumor burden detection on confirmatory biopsies. Moreover, MRI-ultrasound fusion biopsy detected clinical significant cancer in AS patients with far fewer cores while rarely missing Gleason 7 cancer.^{36,37} In the included studies, only Stamatakis *et al.* used MRI-ultrasound-guided targeted biopsy. The AS protocols and nationality of included population had no significance on the results of subgroup evaluation. It should also be noted that the sensitivity in the year after 2012 category and the specificity in the MP-MRI category is better than the opposite category, with $P=0.06$ and $P=0.07$, respectively, which were approaching 0.05. As 1.5 T MRI examinations were performed before or in 2012, and 3 T MRI was almost used after 2012 in the included studies, we found no differences in the diagnostic performance between 1.5 and 3 T MRI studies, and that between MP-MRI and non MP-MRI studies. This suggests that technological improvements in the past decade

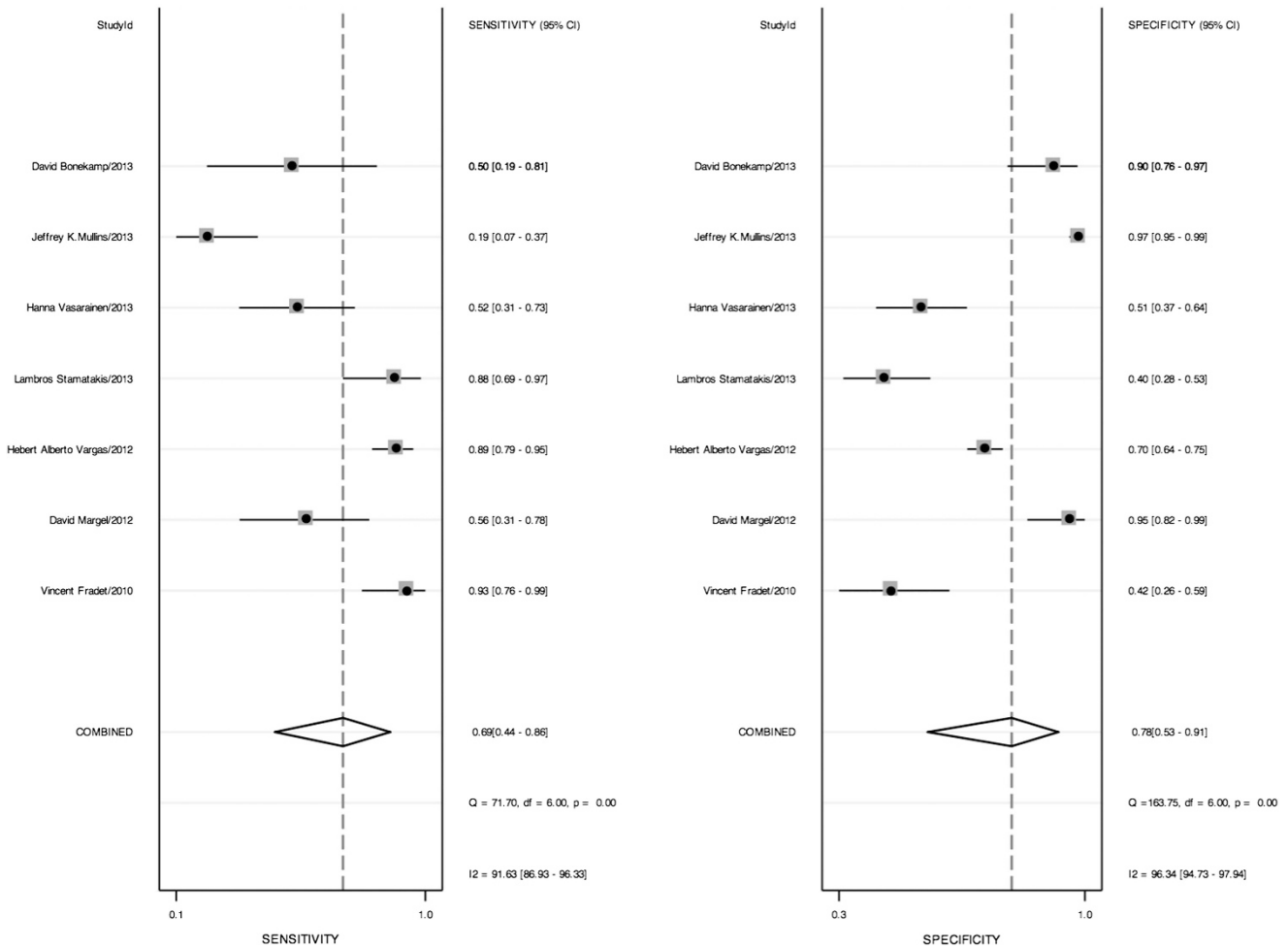


Figure 4. Forest plot showing the sensitivity and specificity of magnetic resonance imaging in the diagnosis of prostate cancer reclassification among active surveillance candidates.

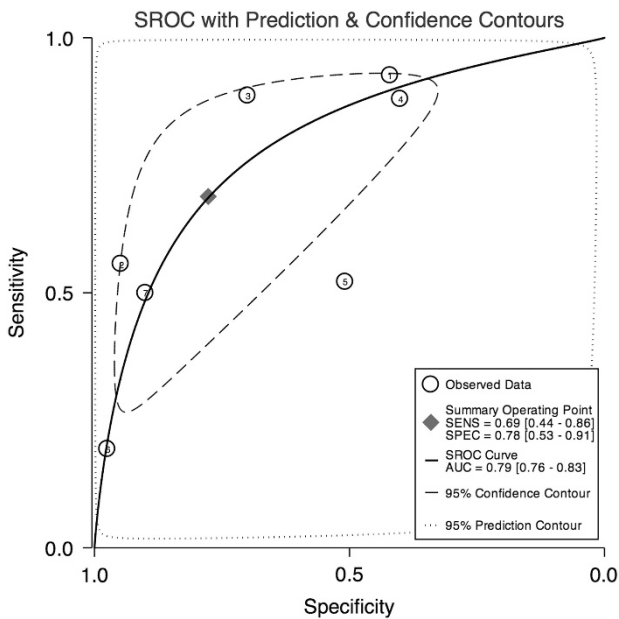


Figure 5. Summary receiver operating characteristic (SROC) curve for assessment of the diagnostic accuracy of magnetic resonance imaging in the diagnosis of prostate cancer reclassification among active surveillance candidates.

may not have significantly affected MRI sequences. 3 T MRI provides prostate images that are natural in shape and that have comparable image quality to those obtained at 1.5 T with an ERC, but not superior diagnostic performance, which suggests that an opportunity exists for improving technical aspects of the 3-T prostate MRI.³⁸ Although there is currently great interest in the use of MP-MRI sequences such as T2-weighted, echo planar, diffusion-weighted imaging and dynamic contrast-enhanced MRI, diffusion-weighted imaging and dynamic contrast-enhanced-MRI have only been clinically available for a few years. Any single technique would lack the sensitivity to detect PCa before AS, since most patients harbor low-volume cancer. Further studies are needed to assess the incremental value of MP-MRI.

The most important finding of our study is that MRI may reveal an unrecognized significant lesion in 33.27% of patients, and biopsy of these areas reclassified 14.59% of cases as no longer fulfilling the criteria for AS. In addition, when no suspicious disease progression (66.34%) was identified on MRI, the chance of reclassification on repeat biopsy was extremely low at 6.13%.

Several earlier studies examined MRI in AS candidates with low-risk PCa. Cabrera *et al.*³⁹ found no association among clinical stage, Gleason, PSA and apparent tumor on endorectal MRI. In another retrospective analysis, van As *et al.*⁴⁰ pointed out that low apparent diffusion coefficient was associated with adverse histology on repeat biopsy. Tumor apparent diffusion coefficient

Table 4. Meta-regression analysis of diagnostic accuracy and summary of diagnostic estimates of subgroup analysis

Subgroup/co-variant	Coefficient	s.e.	P-value	RDOR	Sensitivity, 95% CI	Specificity, 95% CI
Quality	0.605	0.9813	0.5709	1.83		
AS protocols	-0.192	1.0757	0.8667	0.82		
Very low-risk PCa					0.54 (0.17–0.91)	0.86 (0.66–1.00)
Not very low-risk PCa					0.77 (0.56–0.99)	0.69 (0.39–0.99)
P-value					0.28	0.28
Year	-1.696	0.6862	0.0689	0.18		
≤ 2012					0.51 (0.23–0.79)	0.79 (0.54–1.00)
> 2012					0.84 (0.67–1.00)	0.77 (0.45–1.00)
P-value					0.06	0.74
Country	1.009	0.9404	0.3438	2.74		
USA					0.75 (0.52–0.97)	0.77 (0.54–1.00)
Not USA					0.53 (0.07–0.99)	0.77 (0.40–1.00)
P-value					0.32	0.79
MP-MRI	0.466	1.1892	0.7151	1.59		
MP-MRI					0.55 (0.25–0.85)	0.88 (0.76–1.00)
Not MP-MRI					0.83 (0.63–1.00)	0.55 (0.21–0.89)
P-value					0.14	0.07
Endorectal coils	2.334	0.6055	0.0182	10.32		
With ERCs					0.74 (0.52–0.97)	0.81 (0.64–0.99)
Without ERCs					0.52 (-0.18–1.00)	0.51 (-0.18–1.00)
P-value					0.32	0.15
Biopsy	0.063	1.6952	0.9722	1.06		
Targeted biopsy					0.80 (0.66–0.94)	0.63 (0.42–0.83)
Not targeted biopsy					0.31 (0.01–0.61)	0.95 (0.89–1.00)
P-value					0.02	< 0.01

Abbreviations: AS, active surveillance; CI, confidence interval; ERC, endorectal coil; MP-MRI, multiparametric magnetic resonance imaging; PCa, prostate cancer; RDOR, relative diagnostic odds ratio.

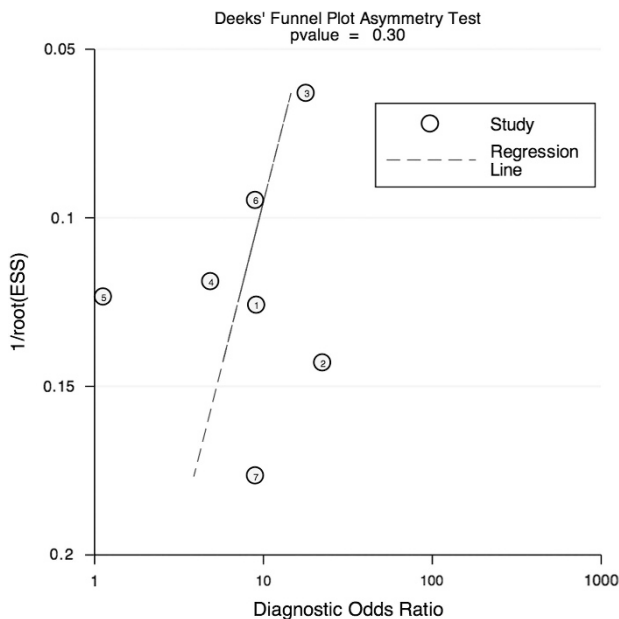


Figure 6. Deeks' funnel plots for publication bias.

correlated with maximum core involvement, the percent of positive cores, initial PSA and the free-to-total PSA ratio.

Limitation should also be considered when we draw conclusion. First, as we have only searched the studies on PubMed, studies on other databases and unpublished data may not be included in our meta-analysis. Second, the sample sizes of some selected studies are small, which can involve overestimating the diagnostic accuracy, and thus subgroup analysis is limited by the restricted original data. Third, variability in imaging techniques, MRI criterion,

reader experience and biopsy techniques are possible confounders. For this reason, the bivariate mixed-effects binary regression modeling was used, as it provides more conservative estimates. Last but not the least, although the funnel plot does not show this meta-analysis as having publication bias, and the number of the studies is small, the influence of publication bias could not be completely excluded.

In summary, this study presents a cumulative analysis of almost 1028 patients from seven studies and demonstrated the value of MRI as a predictor of PCa reclassification among AS candidates. MRI shows a very high NPV for the intermediate end point of disease upstaging. Therefore, favorable MRI findings on high-quality MRI, that is, MP-MRI, may be used for selection and follow-up of AS candidates and might refrain the need for repeat biopsies. However, the PPV of MRI for higher-risk disease seems to be considerably lower in the low-risk PCa patients under AS and may be caused by reporter bias with more false positives in AS cohorts who are known to have PCa. The reported range of percentages of disease progression identified in AS candidates is very wide (19–93%). All these indicate both that a negative MRI in an AS candidate is supportive of keeping him in AS status and that the finding of an MR index lesion, especially that greater than 10 mm, has a potential to become a marker for patients who require increased attention to allow targeted biopsies of index lesions.

CONCLUSION

MRI, especially MP-MRI, has a moderate diagnostic accuracy as a significant predictor of disease reclassification among AS candidates. MRI of the prostate was found to have a high NPV and specificity for the prediction of biopsy reclassification upon clinical follow-up, which suggests that negative prostate MRI findings may support a patient remaining under AS. Although the PPV and sensitivity for the prediction were relatively low, the presence of a suspicious lesion > 10 mm lesion may suggest an increased risk for disease progression.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

- Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T et al. Cancer statistics, 2008. *CA Cancer J Clin* 2008; **58**: 71–96.
- Sanda MG, Dunn RL, Michalski J, Sandler HM, Northouse L, Hembroff L et al. Quality of life and satisfaction with outcome among prostate cancer survivors. *N Engl J Med* 2008; **358**: 1250–1261.
- Ukimura O, Faber K, Gill IS. Intraprostatic targeting. *Curr Opin Urol* 2012; **22**: 97–103.
- Warlick C, Trock BJ, Landis P, Epstein JI, Carter HB. Delayed versus immediate surgical intervention and prostate cancer outcome. *J Natl Cancer Inst* 2006; **98**: 355–357.
- Klotz L, Zhang L, Lam A, Nam R, Mamedov A, Loblaw A. Clinical results of long-term follow-up of a large, active surveillance cohort with localized prostate cancer. *J Clin Oncol* 2010; **28**: 126–131.
- NCCN Guidelines and Clinical Resources. Updated Prostate Cancer Guidelines, vs 2.2014. National Comprehensive Cancer Network. Available at http://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf. Accessed April 2014.
- Porten SP, Whitson JM, Cowan JE, Cooperberg MR, Shinohara K, Perez N et al. Changes in prostate cancer grade on serial biopsy in men undergoing active surveillance. *J Clin Oncol* 2011; **29**: 2795–2800.
- Tosoian JJ, Trock BJ, Landis P, Feng Z, Epstein JI, Partin AW et al. Active surveillance program for prostate cancer: an update of the Johns Hopkins experience. *J Clin Oncol* 2011; **29**: 2185–2190.
- Ross AE, Loeb S, Landis P, Partin AW, Epstein JI, Kettermann A et al. Prostate-specific antigen kinetics during follow-up are an unreliable trigger for intervention in a prostate cancer surveillance program. *J Clin Oncol* 2010; **28**: 2810–2816.
- Loeb S, Carter HB, Berndt SI, Ricker W, Schaeffer EM. Complications after prostate biopsy: data from SEER-Medicare. *J Urol* 2011; **186**: 1830–1834.
- Fujita K, Landis P, McNeil BK, Pavlovich CP. Serial prostate biopsies are associated with an increased risk of erectile dysfunction in men with prostate cancer on active surveillance. *J Urol* 2009; **182**: 2664–2669.
- Zhang J, Hricak H, Shukla-Dave A, Akin O, Ishill NM, Carlino LJ et al. Clinical stage T1c prostate cancer: evaluation with endorectal MR imaging and MR spectroscopic imaging. *Radiology* 2009; **253**: 425–434.
- Shukla-Dave A, Hricak H, Kattan MW, Pucar D, Kuroiwa K, Chen HN et al. The utility of magnetic resonance imaging and spectroscopy for predicting insignificant prostate cancer: an initial analysis. *BJU Int* 2007; **99**: 786–793.
- Delongchamps NB, Beuvon F, Eiss D, Flam T, Muradyan N, Zerbib M et al. Multiparametric MRI is helpful to predict tumor focality, stage, and size in patients diagnosed with unilateral low-risk prostate cancer. *Prostate Cancer Prostatic Dis* 2011; **14**: 232–237.
- Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; **3**: 25.
- Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB et al. QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; **155**: 529–536.
- Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003; **56**: 1129–1135.
- Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001; **323**: 157–162.
- Tang BM, Eslick GD, Craig JC, McLean AS. Accuracy of procalcitonin for sepsis diagnosis in critically ill patients: systematic review and meta-analysis. *Lancet Infect Dis* 2007; **7**: 210–217.
- Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med* 1993; **12**: 1293–1316.
- Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med* 2002; **21**: 1525–1537.
- Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001; **20**: 2865–2884.
- Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005; **58**: 882–893.
- Ziakas PD, Poulou LS, Voulgarelis M, Thanos L. The Gordian knot of interim 18-fluorodeoxyglucose positron emission tomography for Hodgkin lymphoma: a meta-analysis and commentary on published studies. *Leuk Lymphoma* 2012; **53**: 2166–2174.
- Dwamena BA. Evidence-based radiology: step 3—diagnostic systematic review and meta-analysis (critical appraisal). *Semin Roentgenol* 2009; **44**: 170–179.
- Vargas HA, Akin O, Afaq A, Goldman D, Zheng J, Moskowitz CS et al. Magnetic resonance imaging for predicting prostate biopsy findings in patients considered for active surveillance of clinically low risk prostate cancer. *J Urol* 2012; **188**: 1732–1738.
- Fradet V, Kurhanewicz J, Cowan JE, Karl A, Coakley FV, Shinohara K et al. Prostate cancer managed with active surveillance: role of anatomic MR imaging and MR spectroscopic imaging. *Radiology* 2010; **256**: 176–183.
- Stamatakis L, Siddiqui MM, Logan J, Rais-Bahrami S, Walton-Diaz A et al. Accuracy of multiparametric magnetic resonance imaging in confirming eligibility for active surveillance for men with prostate cancer. *Cancer* 2013; **119**: 3359–3366.
- Vasarainen H, Lahdesuo K, Savolainen R, Ruutu M, Taari K, Rannikko A. Diffusion-weighted magnetic resonance imaging in prostate cancer patients on active surveillance one year after diagnosis and before repeat biopsy. *Scand J Urol* 2013; **47**: 456–461.
- Margel D, Yap SA, Lawrentschuk N, Klotz L, Haider M, Hersey K et al. Impact of multiparametric endorectal coil prostate magnetic resonance imaging on disease reclassification among active surveillance candidates: a prospective cohort study. *J Urol* 2012; **187**: 1247–1252.
- Mullins JK, Bonekamp D, Landis P, Begum H, Partin AW, Epstein JI et al. Multiparametric magnetic resonance imaging findings in men with low-risk prostate cancer followed using active surveillance. *BJU Int* 2013; **111**: 1037–1045.
- Bonekamp D, Bonekamp S, Mullins JK, Epstein JI, Carter HB, Macura KJ. Multiparametric magnetic resonance imaging characterization of prostate lesions in the active surveillance population: incremental value of magnetic resonance imaging for prediction of disease reclassification. *J Comput Assist Tomogr* 2013; **37**: 948–956.
- Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* 2006; **6**: 31.
- Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988; **240**: 1285–1293.
- Siadat MS, Shu J. Proportional odds ratio model for comparison of diagnostic tests in meta-analysis. *BMC Med Res Methodol* 2004; **4**: 27.
- Siddiqui MM, Rais-Bahrami S, Truong H, Stamatakis L, Vourganti S, Nix J et al. Magnetic resonance imaging/ultrasound-fusion biopsy significantly upgrades prostate cancer versus systematic 12-core transrectal ultrasound biopsy. *Eur Urol* 2013; **64**: 713–719.
- Da Rosa MR, Milot L, Sugar L, Vesprini D, Chung H, Loblaw A et al. A prospective comparison of MRI-US fused targeted biopsy versus systematic ultrasound-guided biopsy for detecting clinically significant prostate cancer in patients on active surveillance. *J Magn Reson Imaging* 2015; **41**: 220–225.
- Shah ZK, Elias SN, Abaza R, Zynger DL, DeRenne LA, Knopp MV et al. Performance comparison of 1.5-T endorectal coil MRI with 3.0-T nonendorectal coil MRI in patients with prostate cancer. *Acad Radiol* 2015; **22**: 467–474.
- Cabrera AR, Coakley FV, Westphalen AC, Lu Y, Zhao S, Shinohara K et al. Prostate cancer: is inapparent tumor at endorectal MR and MR spectroscopic imaging a favorable prognostic finding in patients who select active surveillance? *Radiology* 2008; **247**: 444–450.
- van As NJ, de Souza NM, Riches SF, Morgan VA, Sohaib SA, Dearnaley DP et al. A study of diffusion-weighted magnetic resonance imaging in men with untreated localised prostate cancer on active surveillance. *Eur Urol* 2008; **56**: 981–987.