

# The utility of quantitative ADC values for differentiating high-risk from low-risk prostate cancer: a systematic review and meta-analysis

Hiram Shaish <sup>1</sup>, Stella K. Kang,<sup>1,2</sup> Andrew B. Rosenkrantz<sup>1</sup>

<sup>1</sup>Department of Radiology, Center for Biomedical Imaging, NYU Langone Medical Center, 660 First Avenue, New York, NY 10016, USA

<sup>2</sup>Department of Population Health, NYU Langone Medical Center, 660 First Avenue, New York, NY 10016, USA

## Abstract

**Purpose:** The purpose of the study is to perform a meta-analysis of studies investigating the diagnostic performance of apparent diffusion coefficient (ADC) values in separating high-risk from low-risk prostate cancer (PCa).

**Methods:** MEDLINE and EMBASE databases were searched in December 2015 for studies reporting diagnostic performance of ADC values for discriminating high-risk from low-risk PCa and providing sufficient data to construct  $2 \times 2$  contingency tables. Diagnostic performance was quantitatively pooled using a bivariate random-effects model including subgroup analysis and assessment of study heterogeneity and methodological quality.

**Results:** 13 studies were included, providing 1107 tumor foci in 705 patients. Heterogeneity among studies was moderate ( $\tau^2 = 0.222$ ). Overall sensitivity was 76.9% (95% CI 68.6–83.6%); overall specificity was 77.0% (95% CI 69.9–82.8%); and summary AUC was 0.67. Inverse correlation between sensitivity and specificity ( $\rho = -0.58$ ) indicated interstudy heterogeneity was partly due to variation in threshold for test positivity. Primary biases were readers' knowledge of Gleason score during ADC measurement, lack of prespecified ADC thresholds, and lack of prostatectomy as reference in some studies. Higher sensitivity was seen in studies published within the past 2 years and studies not using  $b$  value of at least 2000; higher specificity was associated with involvement of one, rather than two, readers measuring ADC. Field strength, coil selection, and advanced diffusion metrics did not significantly impact diagnostic performance.

**Conclusion:** ADC values show moderate accuracy in separating high-risk from low-risk PCa, although important biases may overestimate performance and unexplained sources of heterogeneity likely exist. Further studies using a standardized methodology and addressing identified weaknesses may help guide the use of ADC values for clinical decision-making.

**Key words:** Prostate cancer—MRI—Diffusion—Apparent diffusion coefficient—Meta-analysis

While prostate cancer (PCa) is the most commonly diagnosed non-cutaneous cancer in men in the United States, the large majority of patients die of other causes [1]. Indeed, it is estimated that only 25% of PCa cases would be detected clinically without prostate specific antigen (PSA) screening [2]. Given the high prevalence and relative indolence of most cases of PCa, accurate risk stratification is of paramount importance in order to avoid overtreatment. This need is particularly critical considering the substantial morbidity associated with radical prostatectomy (RP), including high rates of impotence and incontinence in some studies [3]. While the Gleason score (GS) determined at RP is the gold standard for predicting mortality following RP [4], methods are needed to reliably determine the GS prior to RP in order to provide an accurate prognostic assessment and select an appropriate management regimen, for instance between RP and active surveillance. Current alternative risk stratification tools include a combination of laboratory results, clinical findings, and biopsy GS, although these have shown moderate accuracy at best [5–7]. Furthermore, the central role of biopsy in such prognostic schemes necessitates a degree of invasiveness,

**Electronic supplementary material** The online version of this article (doi:10.1007/s00261-016-0848-y) contains supplementary material, which is available to authorized users.

Correspondence to: Hiram Shaish; email: hs2926@cumc.columbia.edu

with associated cost, patient discomfort, and risk of complications.

Diffusion-weighted imaging (DWI) is an MRI technique that has been established to improve tumor detection and localization [8]. Furthermore, an inverse correlation between apparent diffusion coefficient (ADC) values derived from DWI and GS has been well described [9–23]. This association suggests that quantitative ADC values could potentially serve as a prognostic biomarker for PCa treatment selection, and improve upon both the suboptimal accuracy and invasiveness of current risk stratification strategies. However, the extent of the relationship between ADC values and GS has varied between studies, and a greater understanding of this association is warranted in view of the growing interest in both the urologic and radiologic communities in applying ADC values in clinical settings [24]. While past meta-analyses have explored the role of qualitative DWI for tumor detection, no meta-analysis to our knowledge has examined the use of quantitative ADC values for risk stratification. Therefore, the purpose of this study was to determine the performance of quantitative ADC values for separating high-risk from low-risk PCa, as well as to identify factors influencing this performance.

## Methods

### *Literature search*

A single radiologist (HS, a fellow in abdominal MRI) performed searches of the MEDLINE and EMBASE databases to identify relevant articles. The “PICO” question to be addressed by the review was [25] (P) Patients with prostate cancer; (I) Intervention—quantitative ADC measurements obtained from diffusion-weighted MRI; (C) Comparison—none; (O) Outcome/Target Condition—presence of high- or low-risk tumor based on histopathologic findings from either prostate biopsy or radical prostatectomy serving as reference standard. The searches were conducted on December 11th, 2015. The specific search terms are provided in the Appendix in the Electronic Supplementary Material. The study protocol was not registered.

### *Study selection*

The retrieved abstracts were reviewed by the previously noted investigator (HS). Full texts were retrieved for those articles that represented original research and reported diagnostic accuracy of ADC values for separating high-risk from low-risk prostate cancer. To be included in the meta-analysis, the full-text had to then provide sufficient data for constructing a  $2 \times 2$  contingency table. If multiple eligible studies reported overlapping patient cohorts, then the larger study was included. For studies lacking data for construction of the  $2 \times 2$

table but otherwise meeting inclusion criteria, efforts were made to contact the study authors and obtain these data. Articles in non-English languages were translated to English to allow for further assessment.

### *Data extraction*

The included articles were reviewed by two investigators (HS and AR; AR is a fellowship-trained body radiologist with 8 years of experience). The relevant data were independently extracted and differences resolved by consensus. The extracted study data included covariates to be analyzed in meta-analysis, all represented in binary form, as well as other study characteristics.

The selected covariates included the following: Year of publication (classified as before or after January 1, 2014), MRI field strength (classified as 3T for all patients vs. other), GS threshold for differentiating low-risk from high-risk prostate cancer in the study's  $2 \times 2$  data, use of an endorectal coil, use of RP as reference standard in all patients, whether all patients had a known diagnosis of prostate cancer at the time of MRI, number of tumor foci per patient contributing to  $2 \times 2$  data (classified as 1, vs. more than 1 tumor focus, per patient), ADC metric selected for the  $2 \times 2$  data (classified as mean ADC vs. other metric, regardless of inclusion of  $b_0$  in ADC computation), maximal  $b$  value used for the computation of ADC metric within  $2 \times 2$  data (classified as whether or not at least  $2000\text{s/mm}^2$ ), prostate zone studied (classified as whole gland vs. peripheral zone), number of readers placing regions of interest (classified as 2 readers vs. 1 reader), reader experience (classified as whether at least one reader had at least 5 years of experience), and use of anti-peristaltic agent (classified as used in some or all patients vs. not used in any patients).

Additional extracted study characteristics included the country of publication, number of overall patients with tumors contributing to the  $2 \times 2$  contingency table, total number of tumor foci evaluated in the  $2 \times 2$  table, patient age (as reported in the article), and PSA (as reported in the article). In studies involving more than one reader, the reported interreader variability was also recorded, when available.

For each study, a  $2 \times 2$  contingency table was constructed using the extracted data for the performance of ADC values for discriminating high-risk from low-risk prostate cancer. If multiple diagnostic accuracies were reported for ADC values derived from different  $b$  value combinations or for different ADC metrics, then the  $b$  value combination or ADC metric which yielded the highest diagnostic performance (defined as the average of sensitivity and specificity) was selected for inclusion in the meta-analysis. Diffusional kurtosis metrics, when provided in addition to conventional ADC metrics, were not selected. If diagnostic accuracy was reported for separating tumors of varying combinations of GS, then the data

for separating GS 3 + 3 tumor from higher GS tumors were selected. In studies in which diagnostic accuracies were reported for multiple observers, the raw  $2 \times 2$  data of the observers were averaged. If the provided sensitivities and specificities did not yield whole numbers for the  $2 \times 2$  data, then attempts were made to contact the authors for clarification, and the data were fit to the published sensitivities and specificities when unsuccessful. Additional comments regarding derivation of the  $2 \times 2$  data for included studies are detailed in Table 1. The end result was a single  $2 \times 2$  table per study for the meta-analysis.

### Methodological quality assessment

Overall methodological quality and likelihood of bias of the articles included in the meta-analysis were assessed

for risk of bias by using the revised tool for the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) [26]. The questions evaluated for each study are detailed in Table 2. A maximum of 6 months was considered as an acceptable time interval between MRI and biopsy, given the very slow growth rate of PCa with a doubling time of 2 years [27, 28].

Given the expectation of heterogeneity among studies of diagnostic accuracy, the inability of conventional tests for publication bias to handle threshold effects between sensitivity and specificity, as well as uncertainty regarding whether publication bias in fact exists for studies of diagnostic accuracy, tests for publication bias, including the Deeks method, are not recommended for meta-analyses of diagnostic tests and were therefore not performed [29, 30].

**Table 1.**  $2 \times 2$  data of included studies separating high-risk from low-risk prostate cancer with ADC values as well as comments regarding derivation of the  $2 \times 2$  data during full-text evaluation

Study	TP	FP	FN	TN	Comment
Desouza et al. [9]	10	1.3	8	24.7	Used sensitivity of 55% and specificity of 95%, which maximized the average of sensitivity and specificity among provided thresholds. $2 \times 2$ data were fitted to provided sensitivities and specificities
Nagarajan et al. [10]	19	7	4	15	$2 \times 2$ data were provided by the study authors
Rosenkrantz et al. [11]	26	13	25	57	
Kitajima et al. [12]	28	5	14	24	
Bae et al. [13]	9	2	2	29	Used mean ADC based on a $b$ value of 1000 $s/mm^2$ , which yielded the highest accuracy
Itatani et al. [14]	63.5	7	7.5	11	Data represent average of 2 readers
Itatani et al. [15]	38	1	16	6	Used data for ADC normalized to saline, which yielded the highest accuracy. Data represent average of 2 readers
Lebovici et al. [16]	68.5	18	8.5	33	Used ADC normalized to peripheral zone, which yielded highest accuracy. Data represent average of 2 readers
Nowak et al. [17]	58.8	14.6	6.2	24.4	Used sensitivity of 90.5% and specificity of 62.5%, which maximized the average of sensitivity and specificity among provided thresholds
Tamada et al. [18]	46	3	10	10	$2 \times 2$ data were fitted to provided sensitivities and specificities
Min et al. [19]	108	4	18	25	
Roethke et al. [20]	27	1	16	11	Used data for a $b$ value of 2000 $s/mm^2$ which yielded highest accuracy
Wang et al. [21]	54	14	23	35	Used data based on the ADC 10th percentile, which yielded highest accuracy among ADC metrics

**Table 2.** QUADAS-2 questions evaluated for included studies

#### Patient selection

1. Was a consecutive or random sample of patients enrolled?
2. Was a case-control design avoided?
3. Did the study avoid inappropriate exclusions?

#### Index test

4. Were the index test results interpreted without knowledge of the results of the reference standard? (Availability of pathologic findings from biopsy or radical prostatectomy was considered a source of bias for this question.)
5. If a threshold was used, was it prespecified?

#### Reference standard

6. Is the reference standard likely to correctly classify the target condition? (any reference standard other than radical prostatectomy was considered unlikely to correctly grade the PCa.)
7. Were the reference standard results interpreted without knowledge of the results of the index test?

#### Flow and timing

8. Was there an appropriate interval between index test(s) and reference standard? (an interval of over 6 months was considered inappropriate.)
9. Did all patients receive a reference standard?
10. Did all patients receive the same reference standard?
11. Were all patients included in the analysis?

## Statistical analysis

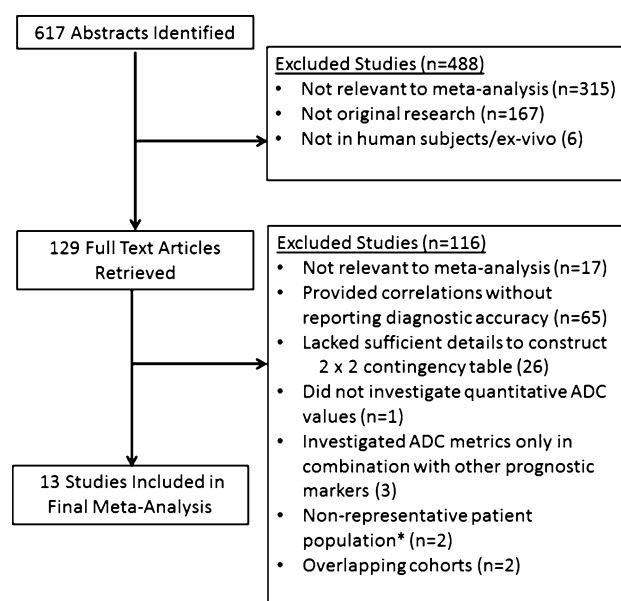
Following computation of the sensitivity and specificity of each individual study for discriminating high-risk from low-risk prostate cancer using an ADC cutoff, the data were pooled and analyzed in sum using the bivariate random-effects model of Reitsma et al. [31]. The sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio across all studies were computed. A summary receiver operating characteristic (sROC) curve was produced and used to calculate the area under the curve (AUC) [32]. A hierarchical summary ROC (HSROC) model was not used given the inability of such a model to allow for separate assessments of study covariates' impact on test sensitivity and specificity, prior studies suggesting similar results in the sROC when using a bivariate random-effects model vs. HSROC model [33], and initial exploratory assessment of our pooled data prior to covariate analyses showing essentially identical sROC curves between the two models (not published).

The presence of heterogeneity between studies was assessed for sensitivity and specificity through the Cochran's  $Q$  test ( $\chi^2 > df, p < 0.1$ : heterogeneity is present) [34] and  $\tau^2$  statistic (quantifies between-study variance when  $>0$ ) [35, 36]. In the case of a direct correlation between the sensitivity and false-positive rate, a threshold effect to explain interstudy heterogeneity was considered likely, whereby apparent differences in performance are explained by a tradeoff between specificity and sensitivity. Sensitivity and specificity were calculated separately for subgroups defined by the prespecified covariates using a bivariate random-effects model. Metaregression was performed of pairwise combinations of selected study characteristics observed to significantly impact study performance at univariable assessment. Statistical analysis was performed using the MADA ("Meta-analysis of diagnostic accuracy") package within the R software environment (R Foundation for Statistical Computing; Vienna, Austria; <http://www.R-project.org>) [37, 38]. A  $p$  value of  $<0.05$  was considered statistically significant.

## Results

### Study selection

The two database searches resulted in 617 abstracts (Fig. 1). 488 of these were excluded at the abstract level (Fig. 1), leaving 129 articles reviewed at the full-text level (Appendix in Electronic Supplementary Material). An additional 116 articles were subsequently excluded (Fig. 1). The remaining 13 studies were included in the meta-analysis [9–21]. One of the included studies [19] was published in Chinese, with an English-language abstract.



**Fig. 1.** Flow chart of study inclusions and exclusions for the meta-analysis. \*Exclusion refers to a study that only enrolled patients with pathologically proven intermediate or high-risk PCa (36), and to a study that only included patients eligible for a strictly defined active surveillance protocol, thereby excluding patients from study based on a broad range of Gleason scores (37).

### Summary of included studies

The 13 included studies are summarized in Tables 3, 4, and 5. Nine studies were published in 2014 or 2015 [13–21]. The MRI field strength was 3T in seven studies [11, 12, 15, 18–21] and 1.5T in the remaining studies. Ten studies classified  $GS \geq 7$  tumors as high risk, while 3 studies used a higher GS threshold to define high risk:  $GS \geq 3 + 4 + 5$  [12],  $GS \geq 4 + 3$  [13], and  $GS \geq 8$  [16]. In 4 studies, an endorectal coil was used [9, 10, 16, 17]. The reference standard was RP for all patients in 7 studies [10, 12–15, 17, 21], biopsy for all patients in 5 studies [9, 11, 16, 19, 20], and a combination of RP and biopsy in 1 study [18]. In 8 studies, all patients had known PCa at the time of MRI [9–12, 16–18, 21]. There was more than one tumor focus per patient contributing to the  $2 \times 2$  data in 9 studies [9–11, 13, 16–19, 21]. In 4 studies, an ADC metric other than mean ADC was selected as providing highest accuracy: normalized ADC in two studies [15, 16], minimum ADC in one study [13], and 10<sup>th</sup> percentile ADC derived from an ADC histogram in one study [21]. Five studies used a  $b$  value greater than or equal to 2000s/mm<sup>2</sup> for calculating ADC [11, 13, 18, 20, 21]. Eight studies included lesions from the whole prostate gland [9, 13–18, 21], while the remaining 5 studies only included peripheral zone lesions. Two readers were involved in placing the ROIs in 10 studies [10–12, 14, 16–21], while one reader was involved in ROI placement in the remaining studies. An expert reader was involved in



**Table 3.** Summary of patient cohorts of included studies

Study	Year published	Country of origin	No. of patients <sup>a</sup>	PSA, ng/ml	Age, years	Study period
Desouza et al. [9]	2008	UK	44	Mean, 10.6	Mean, 66.6	7/2006–12/2006
Nagarajan et al. [10]	2012	USA	41	NA	47–75	2/2007–6/2008
Rosenkrantz et al. [11]	2012	USA	47	Mean, 6.5; 2.8–10.2	Mean, 65; 58–72	12/2010–2/2011
Kitajima et al. [12]	2013	Japan	71	Mean, 9.52; 2.91–49.1	Mean, 66.6; 50–77	1/2010–4/2012
Bae et al. [13]	2014	Japan	29	Median, 6.5; 3.2–18.3	Median, 67; 50–77	11/2009–1/2011
Itatani et al. [14]	2014	Japan	89	Mean, 6.1; 3.1–9.9	Mean, 68.6; 50–80	7/2010–12/2011
Itatani et al. [15]	2014	Japan	58	Mean, 10.9; 4.4–61.1	Mean, 70.7; 58–81	1/2013–2/2014
Lebovici et al. [16]	2014	Romania	22	Mean, 16.3; 3.5–100	Mean, 64.5; 52–75	2/2011–8/2012
Nowak et al. [17]	2014	Germany	66	Mean, 17.5; 1.7–272.4	Mean, 63; 48–73	NA
Tamada et al. [18]	2014	Japan	50	Median, 11.96; 2.35–648	Mean, 70; 53–85	4/2012–3/2013
Min et al. [19]	2015	China	23	NA	Mean, 66; 49–79	8/2013–12/2012
Roethke et al. [20]	2015	Germany	55	Mean, 12.4; 2.7–75	Mean, 67.5; 52–84	1/2012–12/2012
Wang et al. [21]	2015	China	110	Mean, 12.1; 1.5–656	Mean, 69.5; 48–86	8/2013–1/2015

<sup>a</sup> Refers to total number of patients with lesions contributing to 2 × 2 data

**Table 4.** Summary of study design and interpretation scheme of included studies

Study	No. of tumor foci	No. of high-risk tumor foci	GS defined as high risk	Basis of analysis	Reference standard	Number of readers	Independent or consensus interpretations
Desouza et al. [9]	59	18	≥7 <sup>b</sup>	Lesion	TRUS biopsy	1	NA
Nagarajan et al. [10]	45	23	3 + 4	Lesion	RP	2 <sup>a</sup>	NA
Rosenkrantz et al. [11]	121	51	≥7	Lesion	TRUS biopsy	2	Consensus
Kitajima et al. [12]	71	42	≥3 + 4 + 5	Patient	RP	2	Consensus
Bae et al. [13]	42	11	≥4 + 3	Lesion	RP	1	NA
Itatani et al. [14]	89	71	≥7	Patient	RP	2	Independent
Itatani et al. [15]	89	51	≥7	Patient	RP	1	NA
Lebovici et al. [16]	128	77	≥8	Lesion	Transperineal saturation biopsy	2	Independent
Nowak et al. [17]	66	65	≥7	Lesion	RP	2	Consensus
Tamada et al. [18]	69	56	≥3 + 4	Lesion	RP in 17 patients, TRUS biopsy in 33 patients	2	Consensus
Min et al. [19]	155	126	≥7	Lesion	TRUS biopsy	2	Independent
Roethke et al. [20]	55	43	≥7	Patient	Transperineal MRI/US fusion biopsy	2	Consensus
Wang et al. [21]	126	77	≥3 + 4	Lesion	RP	2	Consensus

TRUS, transrectal ultrasound guided; RP, radical prostatectomy

<sup>a</sup> Information obtained through communication with study authors

<sup>b</sup> One patient with GS 3 + 3 was classified as high risk due to PSA, utilizing the NCCN risk groups

**Table 5.** Summary of characteristics of MRI protocol of included studies

Study	Field Strength, T	Endorectal coil	ADC metric	<i>b</i> value ≥2000 s/mm <sup>2</sup>
Desouza et al. [9]	1.5	Yes	ADC <sub>mean</sub> <sup>a</sup>	No
Nagarajan et al. [10]	1.5	Yes	ADC <sub>mean</sub>	No
Rosenkrantz et al. [11]	3.0	No	ADC <sub>mean</sub>	Yes
Kitajima et al. [12]	3.0	No	ADC <sub>mean</sub>	No
Bae et al. [13]	1.5	No	ADC <sub>min</sub>	Yes
Itatani et al. [14]	1.5	No	ADC <sub>mean</sub>	No
Itatani et al. [15]	3.0	No	ADC <sub>normalized</sub>	No
Lebovici et al. [16]	1.5	Yes	ADC <sub>normalized</sub>	No
Nowak et al. [17]	1.5	Yes	ADC <sub>mean</sub>	No
Tamada et al. [18]	3.0	No	ADC <sub>mean</sub>	Yes
Min et al. [19]	3.0	No	ADC <sub>mean</sub>	No
Roethke et al. [20]	3.0	No	ADC <sub>mean</sub>	Yes
Wang et al. [21]	3.0	No	ADC <sub>10th%</sub>	Yes

<sup>a</sup> Mean value was taken from a parametric map of ADC<sub>slow</sub> that excluded *b*<sub>0</sub>

measuring ADC values in five studies [9, 12, 14, 15, 18], while no expert reader was involved in four studies [11, 13, 16, 17]. An antiperistaltic was used prior to prostate MRI in some or all patients in 5 studies [9, 12, 14, 15, 18].

Of the ten studies with two readers, five of these entailed consensus readings. Of the five two-reader studies with independent readings, only one of these reported interobserver variability [15], noting  $\kappa = 0.90$ .

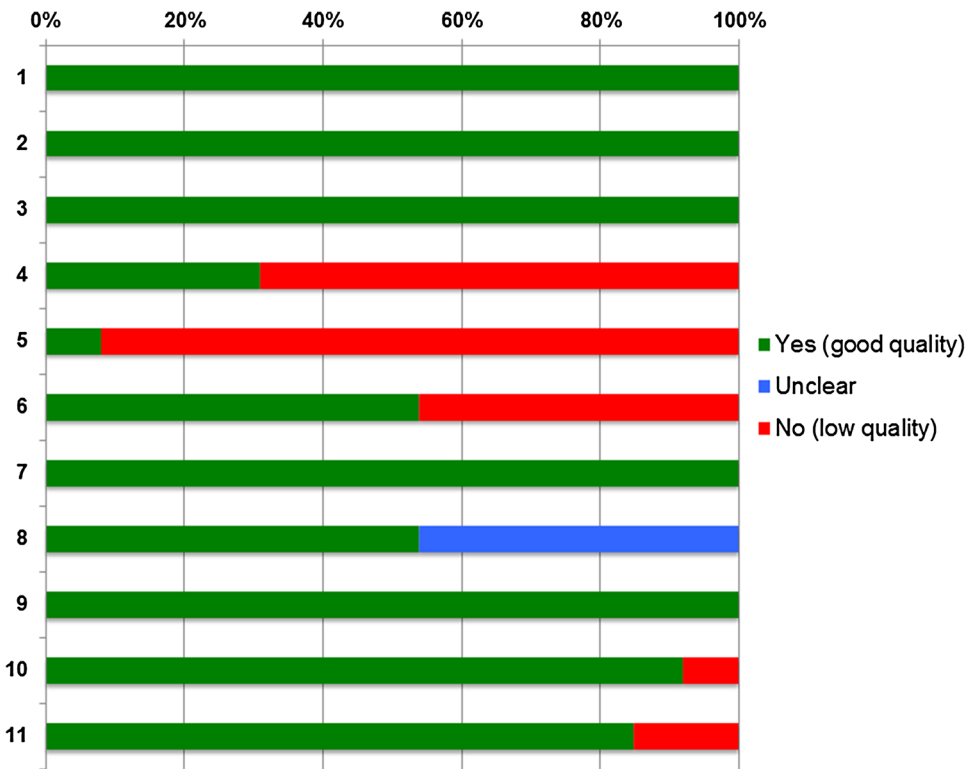


Fig. 2. Risk for bias of studies included within the meta-analysis, evaluated by QUADAS-2 questions. Bar chart demonstrates distribution of responses for the 11 included studies for each question. Green bar = response of “yes” (good quality), blue bar = unclear, and red bar = response of “no” (low quality).

*Quality assessment*

The overall scores for the QUADAS-2 questions are shown in Fig. 2. Scores were very high, indicating low risk of bias, for questions relating to quality of patient selection [100% (13/13) of studies receiving a score “yes”] and quality of the flow and timing of study design [85% (11/13)–100% (13/13)] aside from a score of 54% (7/13) for the question relating to appropriate time interval between MRI and the reference standard. Scores were very low, indicating high risk for bias, for questions relating to the quality of index test interpretation [8% (1/13)–31% (4/13)]. Low performance in this category was attributed to the readers’ access to pathologic findings from either biopsy or radical prostatectomy at the time of ADC measurement as well as lack of a prespecified

ADC threshold. The question relating to the quality of the reference standard received a low to moderate score [54% (7/13)] due to the lack of RP serving as reference standard for all patients in 6 studies.

*Assessment of heterogeneity between studies*

Between-study heterogeneity was quantified using  $\tau^2$  among the 13 studies, with resulting  $\tau^2 = 0.222$ , although Cochran’s Q was not significant (10.81; df = 12,  $p = 0.545$ ). A threshold effect was identified [moderate correlation between sensitivity and false-positive rate ( $r = -0.58$ )], indicating a tradeoff in sensitivity for specificity and accounting for some part of the apparent heterogeneity in test performance among studies.

Study	Sensitivity [95% CI]	Specificity [95% CI]
Rosenkrantz [9]	0.51 (26/51) [0.38, 0.64]	0.81 (57/70) [0.71, 0.89]
Desouza [10]	0.56 (10/18) [0.33, 0.76]	0.96 (24/26) [0.77, 1.00]
Roethke [11]	0.63 (27/43) [0.48, 0.76]	0.92 (11/12) [0.59, 0.99]
Kitajima [12]	0.67 (28/42) [0.51, 0.79]	0.83 (24/29) [0.65, 0.93]
Wang [13]	0.70 (54/77) [0.59, 0.79]	0.71 (35/49) [0.57, 0.82]
Itatani [14]	0.75 (38/54) [0.61, 0.85]	0.86 (6/7) [0.42, 0.98]
Bae [15]	0.82 (9/11) [0.49, 0.95]	0.94 (29/31) [0.78, 0.98]
Tamada [16]	0.82 (46/56) [0.70, 0.90]	0.77 (10/13) [0.48, 0.92]
Nagarajan [17]	0.83 (19/23) [0.62, 0.93]	0.68 (15/22) [0.47, 0.84]
Min [18]	0.86 (108/126) [0.79, 0.91]	0.86 (25/29) [0.69, 0.95]
Lebovici [19]	0.88 (68.5/77) [0.79, 0.94]	0.65 (33/51) [0.51, 0.77]
Itatani [20]	0.89 (63.5/71) [0.79, 0.94]	0.61 (11/18) [0.38, 0.80]
Nowak [21]	0.91 (58.8/65) [0.81, 0.96]	0.62 (24.4/39) [0.46, 0.75]
Summary estimate	0.77 (555.8/713) [0.69, 0.84]	0.77 (305.1/396) [0.70, 0.83]

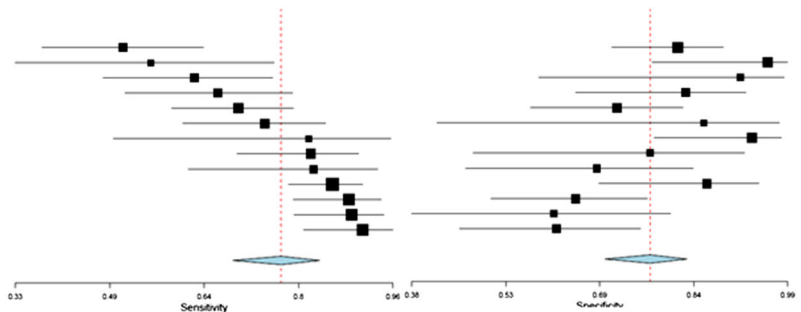
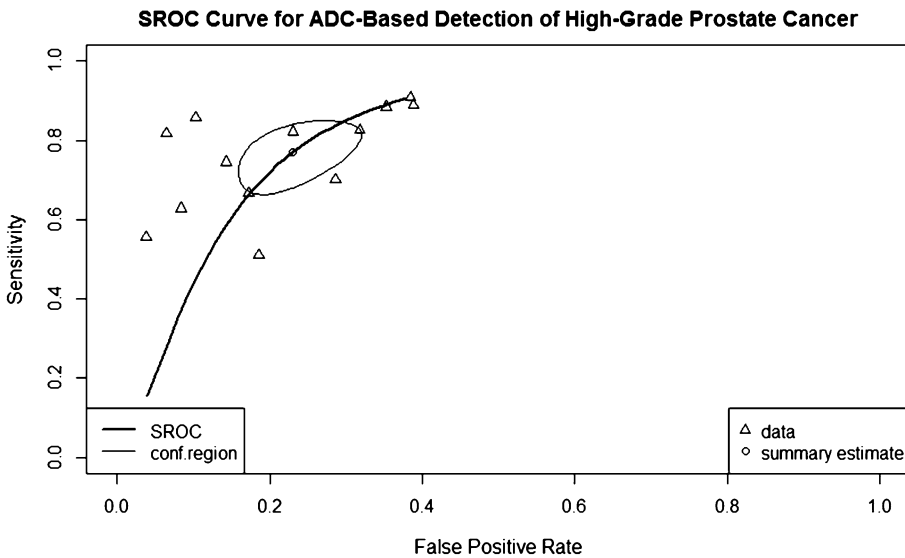


Fig. 3. Forest plots of sensitivity and specificity for quantitative ADC values in separating high-risk from low-risk prostate cancer. Size of squares for individual studies represents

weight of each study. Horizontal lines indicate 95% CIs of individual studies. Studies are ordered by increasing sensitivity.



**Fig. 4.** Summary receiver operating characteristic (sROC) curve of accuracy of quantitative ADC values in separating high-risk from low-risk prostate cancer. The ellipsoid confidence region depicts the 95% confidence region in the sROC space for the summary point estimate of diagnostic performance.

### Overall diagnostic accuracy

There were a total of 1107 tumor foci in 705 patients across the 13 included studies. Sensitivity and specificity were both variable among studies, [range, 51.0% (26/51)–90.8% (58.8/65)] and [range, 61.1% (11/18)–96.2% (24.7/26)], respectively. Figure 3 depicts the Forest plots of sensitivity and specificity for these 13 studies. The summarized sensitivity and specificity of ADC values for separating high-risk from low-risk PCa were 76.9% (95% CI 68.6–83.6%) and 77.0% (95% CI 69.9–82.8%), respectively. The positive likelihood ratio was 3.37 (95% CI 2.66–4.25). The negative likelihood ratio was 0.30 (95% CI 0.22–0.39). Figure 4 shows the sROC curve with AUC of 0.67, indicating moderate performance in discrimination of high-risk from low-risk PCa.

### Subgroup analyses

Table 6 details the results of the subgroup analyses. Sensitivities were significantly higher for studies in which a  $b$  value of 2000s/mm<sup>2</sup> was not included in computation of the ADC values metrics (81.8% vs. 67.1%,  $p = 0.04$ ) and for studies published in 2014 or 2015 (81.3% vs. 63.3%,  $p = 0.02$ ). Specificities were significantly higher for studies in which ADC values were measured by 1 reader rather than by 2 readers (93.3% vs. 73.2%,  $p = 0.005$ ). The remaining analyzed covariates showed no statistically significant difference in terms of sensitivity or specificity ( $p \geq 0.162$ ). At pairwise metaregression, both using ADC values derived from  $b$  values lower than 2000 and publishing in 2014 or 2015 remained as significant independent predictors of increased sensitivity ( $p < 0.001$ ).

## Discussion

Our meta-analysis demonstrated that quantitative ADC values have moderate sensitivity and specificity in sepa-

rating high-risk from low-risk PCa. Nonetheless, both sensitivity and specificity were heterogeneous across the 13 included studies. While studies varied greatly in technical factors, a moderate inverse correlation between reported sensitivity and specificity of individual studies indicates that apparent interstudy heterogeneity was at least partially due to variation along a single ROC curve. Differences in potentially important technical considerations including field strength, coil arrangement, and ADC metric (whether conventional mean ADC or an alternative advanced metric) may also contribute some degree of heterogeneity, though none of these covariates were associated with significant differences in sensitivity or specificity. Nonetheless, our meta-analysis was unable to fully account for all of the observed heterogeneity in test performance across studies, such that additional unexplained sources of heterogeneity likely exist.

The GS determined at RP serves as the single best predictor of mortality in patients undergoing RP for organ-confined PCa [4]. However, the majority of PCa patients have a low-risk tumor with indolent behavior and would be overtreated by RP. The aim in such patients is early and accurate risk stratification in order to avoid RP. To date, no risk stratification tool has been able to predict the GS at RP with sufficient accuracy in order to reliably distinguish between those patients who require RP to prolong cancer-specific survival and those patients who can be managed non-invasively. While the GS determined at systematic biopsy has traditionally served as the mainstay for clinical risk assessments, up to 43% of men with low-grade PCa on standard biopsy will be upgraded on RP [7]. More recently, MRI-targeted biopsy has shown increased accuracy in the determination of tumor grade compared to standard 12 core biopsies [39]. In spite of this improvement, targeted biopsy continues to miss some high-risk lesions. In a large prospective study, Siddiqui et al. demonstrated that

**Table 6.** Estimates of overall sensitivities and specificities for multiple subgroup analyses for separating high-risk from low-risk prostate cancer on MRI (95% CIs listed in parentheses)

Study	Pooled sensitivity	<i>p</i> value	Pooled specificity	<i>p</i> value
Published in 2014 or 2015	81.3% (43.8–94.8%)	<b>0.021</b>	76.6% (39.3–94.3%)	0.547
Published in 2013 or earlier	63.3% (47.3–76.9%)		81.1% (67.3–89.9%)	
MRI performed at 3.0T	71.6% (38.7–91.0%)	0.08	80.8% (54.8–93.6%)	0.216
MRI performed at 1.5T	83.5% (73.4–90.3%)		72.7% (61.3–81.8%)	
Defined high risk as Gleason score >3 + 3 tumor	76.5% (67.7–84.2%)	0.837	77.0% (62.8–83.2%)	0.757
Used a higher GS threshold to define high risk	78.5% (43.3–96.6%)		79.5% (48.7–94.1%)	
Used endorectal coil	82.6% (47.0–94.4%)	0.299	70.5% (42.9–88.3%)	0.162
Did not use endorectal coil	74.4% (64.1–82.7%)		79.9% (72.1–86.0%)	
Radical prostatectomy performed in all patients	79.5% (47.0–94.4%)	0.466	73.5% (41.9–91.4%)	0.233
Radical prostatectomy not performed in all patients	73.8% (60.4–83.9%)		81.6% (71.4–88.8%)	
Known prostate cancer diagnosis at the time of MRI	75.2% (38.1–93.7%)	0.584	73.8% (38.2–92.8%)	0.071
No known prostate cancer diagnosis at the time of MRI	79.5% (65.8–88.7%)		85.7% (74.8–92.4%)	
ADC measurements performed without knowledge of Gleason score	78.6% (44.5–94.4%)	0.807	77.7% (46.9–93.2%)	0.976
ADC measurements performed with knowledge of Gleason score	76.4% (66.0–84.3%)		77.5% (68.3–84.6%)	
More than 1 tumor focus per patient	78.0% (39.3–95.1%)	0.702	76.9% (36.1–95.1%)	0.787
1 tumor focus per patient	74.7% (57.8–86.4%)		79.1% (62.3–89.7%)	
Used an ADC metric other than ADC <sub>mean</sub>	77.6% (44.0–93.9%)	0.908	75.9% (45.2–92.3%)	0.751
Used ADC <sub>mean</sub>	76.7% (66.0–84.7%)		78.3% (69.0–85.4%)	
Maximal <i>b</i> value of greater than or equal to 2000 s/mm <sup>2</sup>	67.1% (36.9–87.8%)	<b>0.044</b>	79.5% (51.0–93.5%)	0.647
Maximal <i>b</i> value less than 2000 s/mm <sup>2</sup>	81.8% (73.6–87.9%)		76.2% (66.0–84.1%)	
Two readers measured ADC values	78.9% (33.0–96.6%)	0.394	73.2% (23.1–96.1%)	<b>0.005</b>
One reader measured ADC values	70.1% (47.4–86.0%)		93.3% (82.6–97.6%)	
Participation of expert <sup>a</sup>	75.5% (26.3–96.4%)	0.782	79.9% (44.9–95.1%)	0.466
No participation of expert reader	78.5% (59.1–90.2%)		73.5% (59.8–83.8%)	
An antiperistaltic was used in some or all patients	75.6% (41.8–93.0%)	0.787	80.1% (51.8–93.8%)	0.594
An anti-peristaltic was not used	77.8% (66.6–86.0%)		76.1% (66.8–83.5%)	
Included lesions from entire prostate	80.3% (48.0–95.9%)	0.250	73.3% (44.3–90.5%)	0.192
Included lesions only from peripheral zone	71.4% (56.6–83.1%)		81.4% (71.7–88.3%)	

Bold values indicate statistical significance at  $p < 0.05$

<sup>a</sup> Expert reader defined as a reader with  $\geq 5$  years experience

fusion biopsy outperformed standard 12 core biopsy in detecting high-risk PCa when using RP as the reference standard (AUC of 0.73 compared to 0.59, respectively) [40]. Other studies have reported a similar trend [41, 42]. While a wide variety of approaches combining results of prostate biopsy with other data elements have also been explored for improved risk stratification, including risk categories incorporating PSA and biopsy findings [5], as well as advanced preoperative nomograms [43], these have also had at best moderate accuracy with AUCs ranging from 0.52 to 0.80 [6].

The previously described prognostic strategies for identifying which patients warrant aggressive treatment

vs. those that can be managed non-invasively all fundamentally rely on information obtained from prostate biopsy. Furthermore, monitoring of patients on surveillance would require serial biopsies to detect a potential change. Our meta-analysis indicates that quantitative ADC values also achieve moderate accuracy in risk stratification, albeit in a completely non-invasive manner. In these regards, based on the available literature, ADC measurements appear to compare favorably with other current invasive schemes for risk stratification of PCa.

Only two of the evaluated study parameters were found to be significantly associated with study sensitivity.



First, sensitivity was significantly higher in more recently published studies, which underscores the potential of continued optimization in enhancing the value of quantitative diffusion metrics in prostate cancer risk assessment. Such optimization may relate to increasing experience within the radiological community in the interpretation of prostate DWI as well as to continued advancements in the acquisition or postprocessing of prostate DWI, for instance involving hardware or software-related factors not captured in our analysis. Furthermore, this observation raises the possibility that the diagnostic performance of DWI for PCa risk assessment may continue to improve with further advances in imaging techniques. Second, sensitivity was significantly lower in studies incorporating a  $b$  value of 2000s/mm<sup>2</sup> in ADC computations. This finding may at first seem surprising given the prior studies that have reported utility from including very high  $b$  value images in clinical assessments of prostate DWI [44, 45]. However, it has also been suggested that despite the value of the high  $b$  value images themselves, these should not be included in ADC map computations given non-mono-exponential diffusion behavior occurring at very high  $b$  values which is not reflected by standard ADC computation techniques, as well as the impact of low SNR at very high  $b$  values on quantitative ADC metrics [46]. Indeed, other studies have failed to confirm benefit of very high  $b$  values for tumor detection when interpretation is based primarily on assessment of the ADC map rather than of the acquired diffusion-weighted images [47–49]. Thus, while there is growing support for very high  $b$  value images, based in part on studies supporting their role for tumor detection, it is possible that the very high  $b$  value images do not have similar added value for non-invasive risk assessment of a detected lesion.

The only study parameter that was observed to be significantly associated with specificity was the number of readers involved in ADC measurement. Specifically, the presence of one, rather than two, reader was associated with significantly increased specificity. The exact explanation for this association is unclear. There was no observed significant association between reader experience and test performance to suggest the presence of a “super-reader” in single-reader studies to account for the finding. Alternatively, we speculate that there may be a tradeoff in sensitivity and specificity when involving multiple readers in selecting areas for region-of-interest measurement, as the presence of 1 reader was also associated with a non-significant decrease in sensitivity. Indeed, the relatively small number of tumor foci in studies with a single reader, rather than 2 readers, performing ADC measurements (144 vs. 687 tumor foci, respectively) may have contributed to the lack of statistical significance for the comparison of sensitivity between these two groups of studies. Further evaluation of the influence of multiple readers is difficult given the

common presence of consensus readings as well as the reporting of interobserver variability in only a single one of the included studies.

While the observed AUC was reassuring, our quality assessment demonstrated several key limitations in the summarized literature that may affect the generalizability of the findings when seeking to apply quantitative ADC values prospectively. Specifically, a number of weaknesses in design were common among the included studies: the availability of pathologic information at the time of ADC measurement, the lack of a prespecified ADC threshold, and (in a smaller number of studies) the lack of radical prostatectomy as reference standard. In addition, some studies that used prostatectomy as reference standard performed radiologic-pathologic correlation on a per-patient rather than a per-lesion basis [12, 14, 15]. As a result, it is possible that the observed accuracy of ADC measurements in the included studies may overestimate the actual test performance of ADC. Thus, prior to prospective application of ADC measurements in clinical management, additional investigations are required to validate the role of ADC values using measurements blinded to pathologic information derived from radical prostatectomy in combination with testing of an a priori determined threshold. Aside from these several important concerns, other key aspects of study quality were overall strong among the included studies.

There are several limitations in our meta-analysis. First, our sample size of 13 studies was small. A lack of availability of  $2 \times 2$  data in many studies examining the role of ADC in PCa risk stratification contributed to this sample size. For example, while a number of studies have explored associations between ADC metrics and prostate cancer aggressiveness, potentially with high-quality study designs including radical prostatectomy as the reference standard, the data of such studies may have largely comprised correlation coefficients or summary AUC values without actual sensitivity and specificity data, thereby precluding formal inclusion in our meta-analysis (Appendix in Electronic Supplementary Material). It is anticipated that additional original data will continue to accumulate as further investigations are performed. Also, we pooled studies utilizing different ADC metrics field strengths, and coil designs. While our meta-analysis showed no significant effect of such technical considerations on accuracy, the study may have been underpowered to detect significant differences in some of the subgroups. In addition, our meta-analysis relies on GS, whether obtained by biopsy or RP, as the reference standard. However, substantial literature demonstrates that determination of the Gleason score itself is variable among pathologists [50–52]. This inherent heterogeneity of the reference standard is a further potential source of variation in performance of ADC metrics not considered by the explored radiological literature. Finally, the ret-

rospective nature of the studies may not reflect the performance of ADC measurements in routine clinical practice.

In conclusion, the available literature supports that quantitative ADC values have moderate accuracy in separating high-risk from low-risk PCa, comparing favorably to that of other currently accepted invasive predictors. Sensitivity was higher in more recent studies and in studies not using a  $b$  value of 2000s/mm<sup>2</sup>, while specificity was higher in studies with a single reader. While heterogeneity in diagnostic performance of ADC across studies could not be fully explained by our meta-analysis, such heterogeneity appeared to represent, at least in part, a tradeoff between sensitivity and specificity. Although the improvement in sensitivity in more recent publications is encouraging regarding the potential future role of quantitative diffusion metrics in prostate cancer risk assessment, methodological heterogeneity exists in numerous technical considerations across studies. Moreover, the available studies may have overestimated the true test performance of ADC due to several identified common methodological biases, including the lack of radical prostatectomy as a reference standard in a number of studies. Thus, additional prospective investigations evaluating the performance of ADC values obtained using a standardized methodology that addresses these particular weaknesses are needed prior to applying ADC values to guide clinical decision-making for PCa.

*Acknowledgments.* We would like to thank Chenchan Huang MD for assistance in English translation of articles.

*Compliance with ethical standards*

*Funding* This study did not receive funding.

*Conflicts of interest* Author A.B.R. receives royalties from Thieme Medical Publishers. Author S. K. K. is supported in part by an AUR GE Radiology Research Academic Fellowship Award. Author H. S. has no conflicts of interests.

*Ethical approval* This meta-analysis did not involve human subjects participants and thus did not require institutional review board approval.

*Informed consent* n/a (not human subjects research).

## References

1. Siegel RL, Miller KD, Jemal A (2015) Cancer statistics, 2015. *CA Cancer J Clin* 65:5–29
2. Etzioni R, Cha R, Feuer EJ, Davidov O (1998) Asymptomatic incidence and duration of prostate cancer. *Am J Epidemiol* 148:775–785
3. Sanda MG, Dunn RL, Michalski J, et al. (2008) Quality of life and satisfaction with outcome among prostate-cancer survivors. *New Engl J Med* 358:1250–1261
4. Epstein JI, Partin AW, Sauvageot J, Walsh PC (1996) Prediction of progression following radical prostatectomy. A multivariate analysis of 721 men with long-term follow-up. *Am J Surg Pathol* 20:286–292
5. Caster JM, Falchook AD, Hendrix LH, Chen RC (2015) Risk of pathologic upgrading or locally advanced disease in early prostate cancer patients based on biopsy Gleason score and PSA: a population-based study of modern patients. *Int J Radiat Oncol Biol Phys* 92:244–251
6. Chun FK, Karakiewicz PI, Briganti A, et al. (2007) A critical appraisal of logistic regression-based nomograms, artificial neural networks, classification and regression-tree models, look-up tables and risk-group stratification models for prostate cancer. *BJU Int* 99:794–800
7. Chun FK, Steuber T, Erbersdobler A, et al. (2006) Development and internal validation of a nomogram predicting the probability of prostate cancer Gleason sum upgrading between biopsy and radical prostatectomy pathology. *Eur Urol* 49:820–826
8. Tan CH, Wei W, Johnson V, Kundra V (2012) Diffusion-weighted MRI in the detection of prostate cancer: meta-analysis. *AJR Am J Roentgenol* 199:822–829
9. deSouza NM, Riches SF, Vanas NJ, et al. (2008) Diffusion-weighted magnetic resonance imaging: a potential non-invasive marker of tumour aggressiveness in localized prostate cancer. *Clin Radiol* 63:774–782
10. Nagarajan R, Margolis D, Raman S, et al. (2012) MR spectroscopic imaging and diffusion-weighted imaging of prostate cancer with Gleason scores. *J Magn Reson Imaging: JMRI* 36:697–703
11. Rosenkrantz AB, Sigmund EE, Johnson G, et al. (2012) Prostate cancer: feasibility and preliminary experience of a diffusional kurtosis model for detection and assessment of aggressiveness of peripheral zone cancer. *Radiology* 264:126–135
12. Kitajima K, Takahashi S, Ueno Y, et al. (2013) Do apparent diffusion coefficient (ADC) values obtained using high b-values with a 3-T MRI correlate better than a transrectal ultrasound (TRUS)-guided biopsy with true Gleason scores obtained from radical prostatectomy specimens for patients with prostate cancer? *Eur J Radiol* 82:1219–1226
13. Bae H, Yoshida S, Matsuoka Y, et al. (2014) Apparent diffusion coefficient value as a biomarker reflecting morphological and biological features of prostate cancer. *Int Urol Nephrol* 46:555–561
14. Itatani R, Namimoto T, Kajihara H, et al. (2014) Triage of low-risk prostate cancer patients with PSA levels 10 ng/ml or less: comparison of apparent diffusion coefficient value and transrectal ultrasound-guided target biopsy. *AJR Am J Roentgenol* 202:1051–1057
15. Itatani R, Namimoto T, Yoshimura A, et al. (2014) Clinical utility of the normalized apparent diffusion coefficient for preoperative evaluation of the aggressiveness of prostate cancer. *Jpn J Radiol* 32:685–691
16. Lebovici A, Sfrangeu SA, Feier D, et al. (2014) Evaluation of the normal-to-diseased apparent diffusion coefficient ratio as an indicator of prostate cancer aggressiveness. *BMC Med Imaging* 14:15
17. Nowak J, Malzahn U, Baur AD, et al. (2014) The value of ADC, T2 signal intensity, and a combination of both parameters to assess Gleason score and primary Gleason grades in patients with known prostate cancer. *Acta radiologica (Stockholm, Sweden: 1987)* 2014
18. Tamada T, Kanomata N, Sone T, et al. (2014) High b value (2,000 s/mm<sup>2</sup>) diffusion-weighted magnetic resonance imaging in prostate cancer at 3 Tesla: comparison with 1,000 s/mm<sup>2</sup> for tumor conspicuity and discrimination of aggressiveness. *PLoS ONE* 9:e96619
19. Min X, Wang L, Feng Z, et al. (2015) Prostate cancer: the correlation between apparent diffusion coefficient values obtained from high resolution diffusion-weighted imaging and Gleason scores. [Chinese]. *Chinese J Radiol (China)* 49:191–194
20. Roethke MC, Kuder TA, Kuru TH, et al. (2015) Evaluation of diffusion kurtosis imaging versus standard diffusion imaging for detection and grading of peripheral zone prostate cancer. *Invest Radiol* 50:483–489
21. Wang Q, Li H, Yan X, et al. (2015) Histogram analysis of diffusion kurtosis magnetic resonance imaging in differentiation of pathologic Gleason grade of prostate cancer. *Urol Oncol* 33:337.e315–324
22. Chang JH, Lim Joon D, Lee ST, et al. (2014) Diffusion-weighted MRI, 11C-choline PET and 18F-fluorodeoxyglucose PET for predicting the Gleason score in prostate carcinoma. *Eur Radiol* 24:715–722

23. Kim TH, Jeong JY, Lee SW, et al. (2015) Diffusion-weighted magnetic resonance imaging for prediction of insignificant prostate cancer in potential candidates for active surveillance. *Eur Radiol* 25:1786–1792
24. Chen L, Liu M, Bao J, et al. (2013) The correlation between apparent diffusion coefficient and tumor cellularity in patients: a meta-analysis. *PLoS ONE* 8:e79008
25. McInnes MD, Bossuyt PM (2015) Pitfalls of systematic reviews and meta-analyses in imaging research. *Radiology* 277:13–21
26. Whiting PF, Rutjes AW, Westwood ME, et al. (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 155:529–536
27. Carter HB, Morrell CH, Pearson JD, et al. (1992) Estimation of prostatic growth using serial prostate-specific antigen measurements in men with and without prostate disease. *Cancer Res* 52:3323–3328
28. Schmid HP, McNeal JE, Stamey TA (1993) Observations on the doubling time of prostate cancer. The use of serial prostate-specific antigen in patients with untreated disease as a measure of increasing cancer volume. *Cancer* 71:2031–2040
29. Leeflang MM, Deeks JJ, Takwoingi Y, Macaskill P (2013) Cochrane diagnostic test accuracy reviews. *Syst Rev* 2:82
30. van Enst WA, Ochodo E, Scholten RJ, Hooft L, Leeflang MM (2014) Investigation of publication bias in meta-analyses of diagnostic test accuracy: a meta-epidemiological study. *BMC Med Res Methodol* 14:70
31. Reitsma JB, Glas AS, Rutjes AW, et al. (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 58:982–990
32. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29–36
33. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA (2007) A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 8:239–251
34. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. *BMJ* 327:557–560
35. DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. *Control Clin Trials* 7:177–188
36. Higgins JP, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. *Stat Med* 21:1539–1558
37. Doebler P (2012) mada: Meta-analysis of diagnostic accuracy. <https://r-forge.r-project.org/projects/mada/>. Accessed 19 Aug 2016
38. Doebler P, Holling H, Bohning D (2012) A mixed model approach to meta-analysis of diagnostic studies with binary test outcome. *Psychol Methods* 17:418–436
39. Siddiqui MM, Rais-Bahrami S, Truong H, et al. (2013) Magnetic resonance imaging/ultrasound-fusion biopsy significantly upgrades prostate cancer versus systematic 12-core transrectal ultrasound biopsy. *Eur Urol* 64:713–719
40. Siddiqui MM, Rais-Bahrami S, Turkbey B, et al. (2015) Comparison of MR/ultrasound fusion-guided biopsy with ultrasound-guided biopsy for the diagnosis of prostate cancer. *JAMA* 313:390–397
41. Le JD, Stephenson S, Brugger M, et al. (2014) Magnetic resonance imaging-ultrasound fusion biopsy for prediction of final prostate pathology. *J Urol* 192:1367–1373
42. Sonn GA, Chang E, Natarajan S, et al. (2014) Value of targeted prostate biopsy using magnetic resonance-ultrasound fusion in men with prior negative biopsy and elevated prostate-specific antigen. *Eur Urol* 65:809–815
43. Shariat SF, Karakiewicz PI, Suardi N, Kattan MW (2008) Comparison of nomograms with other methods for predicting outcomes in prostate cancer: a critical analysis of the literature. *Clin Cancer Res* 14:4400–4407
44. Metens T, Miranda D, Absil J, Matos C (2012) What is the optimal b value in diffusion-weighted MR imaging to depict prostate cancer at 3T? *Eur Radiol* 22:703–709
45. Katahira K, Takahara T, Kwee TC, et al. (2011) Ultra-high-b-value diffusion-weighted MR imaging for the detection of prostate cancer: evaluation in 201 cases with histopathological correlation. *Eur Radiol* 21:188–196
46. Rosenkrantz AB, Padhani AR, Chenevert TL, et al. (2015) Body diffusion kurtosis imaging: basic principles, applications, and considerations for clinical practice. *J Magn Reson Imaging: JMRI* 42:1190–1202
47. Rosenkrantz AB, Hindman N, Lim RP, et al. (2013) Diffusion-weighted imaging of the prostate: comparison of b1000 and b2000 image sets for index lesion detection. *J Magn Reson Imaging: JMRI* 38:694–700
48. Kitajima K, Kaji Y, Kuroda K, Sugimura K (2008) High b-value diffusion-weighted imaging in normal and malignant peripheral zone tissue of the prostate: effect of signal-to-noise ratio. *Magn Reson Med Sci: MRMS* 7:93–99
49. Kim CK, Park BK, Kim B (2010) High-b-value diffusion-weighted imaging at 3 T to detect prostate cancer: comparisons between b values of 1,000 and 2,000 s/mm<sup>2</sup>. *AJR Am J Roentgenol* 194:W33–W37
50. Oyama T, Allsbrook WC Jr, Kurokawa K, et al. (2005) A comparison of interobserver reproducibility of Gleason grading of prostatic carcinoma in Japan and the United States. *Arch Pathol Lab Med* 129(8):1004–1010
51. De la Taille A, Villefond A, Berger N, et al. (2003) Evaluation of the interobserver reproducibility of Gleason grading of prostatic adenocarcinoma using tissue microarrays. *Hum Pathol* 34(5):444–449
52. Allsbrook WC Jr, Mangold KA, Johnson MH, et al. (2001) Interobserver reproducibility of Gleason grading of prostatic carcinoma: urologic pathologists. *Hum Pathol* 32(1):74–80