Title

Accuracy of high b-value diffusion-weighted MRI for prostate cancer detection: a meta-analysis

Authors

Keith C Godley[1], Tom Syer[2], Andoni P Toms[1, 2], Toby O Smith[3], Paul N Malcolm[1,2]

1. Radiology Academy, Norfolk and Norwich University Hospital, Norwich, UK
2. Norwich Medical School, University of East Anglia, Norwich, UK
3. School of Health Sciences, University of East Anglia, Norwich, UK

Abstract

Background: The diagnostic accuracy of diffusion-weighted imaging (DWI) to detect prostate cancer is well-established. DWI provides visual and also quantitative means of detecting tumor, the Apparent Diffusion Coefficient (ADC). Recently higher b-values have been used to improve DWI's diagnostic performance.

Purpose: To determine the diagnostic performance of high b-value DWI at detecting prostate cancer and whether quantifying ADC improves accuracy.

Material and Methods: A comprehensive literature search of published and unpublished databases was performed. Eligible studies had histopathologically proven prostate cancer, DWI sequences using b-values ≥ 1000 s/mm^2, > 10 patients, and data for creating a 2x2 table. Study quality was assessed with QUADAS-2 (Quality Assessment of diagnostic Accuracy Studies). Sensitivity and specificity were calculated and tests for statistical heterogeneity and threshold effect performed. Results were plotted on a summary receiver operating characteristic curve (sROC) and the area under the curve (AUC) determined the diagnostic performance of high b-value DWI.

Results: Ten studies met eligibility criteria with 13 subsets of data available for analysis, including 522 patients. Pooled sensitivity and specificity were 0.59 (95% CI 0.57–0.61) and 0.92 (95% CI 0.91–0.92) respectively and the sROC AUC was 0.92. Subgroup analysis showed a statistically significant (p=0.03) improvement in accuracy when using tumor visual assessment rather than ADC.

Conclusion: High b-value DWI gives good diagnostic performance for prostate cancer detection and visual assessment of tumor diffusion is significantly more accurate than ROI measurements of ADC.

Keywords: Prostate cancer, MRI, diffusion-weighted imaging, DWI, high-b-value, meta-analysis

PROSPERO Registration Number: CRD42015027644

Introduction

Prostate cancer is the most commonly diagnosed cancer, and second-most-common cause of cancer-related death in men (1,2). Magnetic resonance imaging (MRI) is the imaging mainstay of prostate cancer localization, being recommended in men considered for radical treatment following positive trans-rectal ultrasound (TRUS) biopsy, and in high risk patients with a negative biopsy, active surveillance, response to treatment, and recurrence (3,4).

Multi-parametric MRI of the prostate comprises T1-weighted and T2-weighted imaging, with additional techniques such as diffusion-weighted imaging (DWI) and dynamic contrast-enhanced MRI (3). DWI has become a routine part of prostate MRI protocols, as it provides good tumor contrast without exogenous agents.

Applying DWI with multiple diffusion weightings, or b-values, allows the apparent diffusion coefficient (ADC) to be estimated: a parameter known to inversely correlate with tumor aggressiveness (5,6). Clinically, both diffusion-weighted images, typically higher b-values, and ADC maps are assessed to detect tumor, which appears bright on diffusion images and dark on ADC maps.

ADC maps are calculated through monoexponential fits to diffusion signal decays on a voxel-by-voxel basis. Other signal models, have been applied to prostate DWI,

including diffusional kurtosis (7), intravoxel incoherent motion (8) the stretched exponential (9), and VERDICT (10). However, the monoexponential model is the most common, requiring only two b-values for fitting.

Performing monoexponential fitting with higher maximum b-values, b_{max}, can improve contrast-to-noise in the resulting maps, where contrast-to-noise is defined as: (Signal(lesion) – Signal (background))/estimated noise [R1 M2]. This gives better characterization of ADC differences between normal and cancerous prostate, improving tumor detection at the expense of reduced signal-to-noise and increased sensitivity to motion artifacts. At 1.5T, b_{max} of 500 and 1000 s/mm^2 are typically used, but increased signal-to-noise at higher field strengths permits the use of higher b_{max} (11). Guidelines recommend a b_{max} of 800–1000 s/mm^2 (3); but there is no consensus on whether higher b-values ($b_{max} \geq 1000$ s/mm^2) should be used routinely. Many studies have compared high and lower b-value (bmax < 1000 s/mm2) acquisitions, but results have been conflicting (12–17).

The diagnostic accuracy of DWI is well-established, with multiple meta-analyses reporting its diagnostic performance (18–20). There is uncertainty about the usefulness of high b-value DWI, particularly which b-value provides ADC maps with the greatest diagnostic performance. The aim of this study is to determine the diagnostic performance

of high b-value DWI at detecting prostate cancer and whether quantifying ADC improves accuracy.

Material and Methods

This meta-analysis was reported using the preferred reporting items for systematic reviews and meta-analyses outlined in the PRISMA statement (21). The review was registered prior to commencing on PROSPERO (Ref No: CRD42015027644).

Search strategy

A comprehensive systematic literature search was independently performed by two reviewers (KCG, TSy) to identify studies investigating the diagnostic accuracy or performance of high b-value DWI for detecting prostate cancer. A MEDLINE search is presented in Supplementary Table 1. In addition, searches were conducted of EMBASE, and the grey literature/trial registry databases: WHO International Clinical Trials Registry Platform and OpenGrey. Studies were not limited by country of origin, but were limited to those published in English. All searches were from database inception to 1st January 2016.

Eligibility criteria

Retrospective and prospective studies were included if they reported detection of prostate cancer in a pre-treatment population using high b-value DWI of the prostate. Only primary research articles, available as full-text, were accepted; however, review articles were checked for additional primary references. High b-value was defined as $b_{max} \geq 1000$ s/mm^2. Histopathological results as a reference standard (biopsy or radical prostatectomy), sufficient data to calculate true positive (TP), false positive (FP), false negative (FN), and true negative (TN) data were required. If multiple b-values were used, including b < 1000 s/mm^2, the study was only eligible if data with b ≥ 1000 s/mm^2 could be extracted. Studies using high b-value DWI in combination with other diagnostic sequences to detect cancer were excluded.

Study identification

Titles and abstracts from the search results, and the full-text papers for all studies which met or potentially met the eligibility criteria, were independently reviewed by two reviewers (KCG, TSy). Those studies which met the eligibility criteria on full review

were included in the final analysis. Disagreements on inclusion suitability were resolved by consensus (KCG, TSy).

Data extraction

Two reviewers (KCG, TSy) independently extracted the data on a pre-defined template, including: publication year, country of origin, sample size, description of study population (age), study design (prospective, retrospective, or unknown), patient enrollment (consecutive or not), inclusion and exclusion criteria, reasons for exclusions from analysis, and number of experts who assessed and interpreted MRI results. Data were recorded on: blinding of MRI measurements to clinical, biochemical or histopathological results; methods used to determine diagnosis; types of coils; and b-values used. For each study, we recorded the number of true-positive, false-positive, true-negative, and false-negative findings for high b-value DWI in diagnosing prostate cancer. Disagreements in data extraction findings were resolved through discussion or through adjudication with a third reviewer (TSm).

Quality Assessment

Two reviewers (KCG, TSy) independently assessed each included paper's quality using QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) (22). Any

disagreements were resolved through discussion or through adjudication with a third reviewer (TSm).

Statistical analysis

Study heterogeneity was assessed through examination of the data extract table. This indicated broad study homogeneity, meaning a meta-analysis was appropriate. Statistical heterogeneity was assessed using the chi-squared statistic, Q, and the inconsistency, I^2. When $p < 0.10$ and $I^2 > 50\%$, unexplained statistical heterogeneity was evident and diagnostic performance analyses were performed using a random-effects model.

Specificity and sensitivity of each study was calculated using 2 × 2 contingency tables. Pooled sensitivity and specificity and positive and negative likelihood ratios with 95% CIs were calculated. Finally, the specificity and sensitivity were used to calculate a summary receiver operating characteristic (sROC) curve and the area under the curve (AUC).

The threshold effect was assessed visually, by determining whether the sROC curve presented with a ‘shoulder-arm’ shape, and qualitatively using the Spearman correlation coefficient of the logit of sensitivity and the logit of (1-specificity), with p <0.05 indicating the heterogeneity between studies could not be explained by threshold effect.

A meta-regression and subgroup analysis was performed to explore other sources of heterogeneity and how they influence diagnostic performance.

All statistical computations were performed using Meta-DiSc (version 1.4, Javier Zamora) and Review Manager (version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014).

Results

Search results

A summary of the search strategy results is presented in Supplementary Fig. 1. In total, 351 studies were identified from the search results, of which 61 were deemed potentially eligible. After full-text review, 10 studies met the final eligibility criteria and were included in the analysis (16,17,23–30).

Characteristics of included studies

The principle characteristics of the included studies are displayed in Table 1, with imaging and study methods listed in Table 2. From the 10 included studies, 522 patients

were analyzed, with mean (range) age = 64 (43–87) years. The mean and median prostate-specific antigen was 19 and 9.3 ng/mL, respectively.

Three studies were prospective and seven retrospective. Field strengths of 1.5T (23,25,26,28,29) and 3T (16,17,24,27,30) were each used in five studies. Radical prostatectomy specimens were used as the reference standard in six studies (16,17,25,27,29,30), biopsy specimens in three (23,24,26) and one study used a combination (28). The MRI reader was blinded in eight studies (16,17,23,25–28,30), blinding was not known in two. Anti-spasmodic agents, either glucagon or hyoscine butylbromide, were used in five studies (16,17,24,29,30) and their use unknown in the remainder. Five studies (16,17,24,29,30) used b-values > 1000 s/mm^2.

Several methods were used to detect prostate cancer: region-of-interest (ROI) based ADC quantification was used in six studies (23,24,26–29); and visual assessment of lesions using ADC maps was performed in four studies (16,17,25,30), of which two (17,30) used a scale (such as Likert scale), and the other two used a binary cut-off.

All ten studies used the monoexponential function only to estimate ADC. In 3 studies (16,17,27), extracted data were split into subsets. Kim et al. (16) and Koo et al. (17) generated multiple ADC maps: b = (0, 1000), and (0, 2000) s/mm^2 for the former, and b = (0, 300), (0, 700), (0, 1000), and (0, 2000) s/mm^2 for the latter. Rosenkrantz et al. (27) split results from the peripheral zone and the transitional zone. The other studies generated only one set of ADC maps from their DWI data, performing monoexponential fitting to all acquired b-values.

Quality assessment

Study quality assessment is presented in Supplementary Table 2. Fig. 2 demonstrates the QUADAS-2 graphical summary of the studies indicating the proportion of high, low, or unclear risk in each domain. A high risk of bias was demonstrated in the patient selection domain, but overall the quality of the studies included was considered ‘good’.

Diagnostic performance

Diagnostic results of the meta-analyses are presented in Table 3. The results from the individual studies are presented in Supplementary Table 3.

The pooled sensitivity and specificity of high b-value DWI MRI in detecting prostate cancer were 0.59 (95% CI: 0.57–0.61; Fig. 2) and 0.92 (95% CI: 0.91–0.92; Fig. 3), respectively. Sensitivity and specificity heterogeneity tests gave $Q = 435.05$ ($p \ll 0.001$), $I^2 = 97.2\%$ and $Q = 89$ ($p \ll 0.001$), $I^2 = 86.5\%$ respectively, indicating significant statistical heterogeneity between studies.

The pooled positive and negative likelihood ratios for high b-value DWI MRI in detecting prostate cancer were 6.64 (95% CI: 4.9–9.0; Supplementary Fig. 2) and 0.33 (95% CI: 0.2–0.5; Supplementary Fig. 3), respectively. Positive and negative likelihood ratio heterogeneity tests gave $Q = 82.50$ ($p \ll 0.001$), $I^2 = 85.5\%$ and $Q = 517.45$ ($p \ll 0.001$), $I^2 = 97.7\%$, respectively, indicating significant statistical heterogeneity between studies.

Fig. 4 shows the sROC curve of the 10 studies, where AUC = 0.92, indicating ‘good’ diagnostic accuracy (31).

Meta-regression analysis

The ROC curve did not demonstrate a ‘shoulder-arm’ shape (Supplementary Table 4**)** and the Spearman Correlation Coefficient between the logit of sensitivity and the logit of (1-

specificity) was 0.286 (p=0.344), confirming that the threshold effect is not responsible for the variation in accuracy between studies.

Subgroup analysis

Subgroup analysis was based on different study characteristics and perceived sources of bias and applicability uncovered in the QUADAS assessment. Studies at 3T with and without an endorectal coil demonstrated the highest pooled sensitivity of 0.76 (95% CI: 0.71–0.80) and 0.74 (95% CI: 0.71–0.79) respectively. When assessing protocols with a b_{max} > 1000 s/mm^2, the pooled specificity and AUC of the sROC were greater: 0.94 (95% CI: 0.93–0.95) and 0.98, respectively. A statistically significant (p<0.05) improvement was seen using assessment of tumor presence on ADC maps as a visual threshold versus ROI measurements. The diagnostic performance of the subgroup analysis and p-values of the above-mentioned factors and others are demonstrated in Table 3.

Discussion

This analysis indicates that high b-value imaging is a good diagnostic tool for detecting prostate cancer. The results of the threshold method subgroup analysis imply that there is a benefit in using higher b_{max} in a clinical setting. The lesser value of quantitative ADC

thresholding as a tool for detecting tumor is in line with PI-RADS version 2 recommendations (standardized reporting standards for prostate MRI) (32). The evidence on which this analysis was made was graded as 'good' quality using the QUADAS-2 tool.

There have been multiple meta-analyses investigating the diagnostic accuracy of DWI alone or in combination with other imaging techniques (18–20). The pooled sensitivity, specificity, and overall accuracy of our study were 0.58, 0.92 and 0.92 respectively, similar to Jie et al's (19) meta-analysis of DWI alone. This is likely due to overlap of included studies, with nine of the ten included studies featured in their meta-analysis. However, in contrast to Jin et al.'s meta-analysis of all b-values, the sensitivity was lower in our study (0.58 vs. 0.77), but the pooled specificity and AUC were higher (0.92 vs. 0.84 and 0.92 vs. 0.88 respectively) (20). This suggests high b-value imaging may help to rule out significant prostate cancer.

There was significant statistical heterogeneity between the included studies that could not be explained by threshold effect. Subgroup analyses of multiple study parameters were assessed to attempt to explain the heterogeneity. Given the unknown cause of statistical heterogeneity, these findings should be interpreted cautiously.

Improved tumor contrast at high b-values comes at the cost of decreased signal-to-noise (11); however, this can be mitigated through the use of 3T field strength. Most of the diagnostically specific high b-value diagnostic accuracy studies use 3T (12,33–35). The subgroup analysis of field strength demonstrated a trend towards improved accuracy with 3T. The sensitivity results of the 3T group alone are similar and the specificity and accuracy are better than those found in a meta-analysis of accuracy of visual assessment of combined T2 and DWI sequences for prostate cancer detection by Wu et al. (21).

On review of the method of prostate cancer detection, about half of the studies qualitatively assessed the ability of blinded readers to visually detect prostate cancer either by answering a binary question regarding cancer presence or using a probability scale (16,17,25,30). The remaining studies used ROI-based ADC calculations to determine prostate cancer presence or absence, or used a scale of ADC values to predict cancer (23,24,26–29). The diagnostic performance of tumor visual assessment on ADC maps was significantly better than quantitative ADC methods, with visual assessment giving similar results to Wu et al.'s combined T2 and DWI meta-analysis (18), indicating potential value for high b-value imaging clinically.

A potential explanation for the relatively poor performance of quantitative ADC methods is four of the six studies included used biopsy as a reference standard whereas all visual assessment studies used radical prostatectomy specimens. TRUS biopsy as a

reference standard is limited given its poor octant localization, small sampling volume, and substantial number of tumors missed (36,37). Radical prostatectomy studies performed better than biopsy studies, but the difference was not statistically significant.

ADC estimation is influenced by a number of factors, including, but not limited to: noise, fat, and perfusion signals (8,38,39); possible non-Gaussian diffusion (7); and the known diffusion anisotropy of the prostate (40). Because of these confounds, other diffusion models may ultimately prove more appropriate for identifying prostate cancer (41,42) [R1 M3]. Along with the study-specific TE, diffusion gradient duration, δ, and diffusion time, Δ, such factors could substantially influence the sensitivity and specificity of DWI for evaluating prostate cancer. None of the studies reviewed reported δ and Δ; Few researchers (42,43) have considered these factors when applying prostate DWI [R1 M1].

None of the other subgroup analyses demonstrated a significant difference between the groups. In some subgroups a statistical difference would have been difficult to demonstrate given the small numbers of studies. For example, only three studies used biopsy as a reference standard, but despite this, this subgroup analysis provided the second strongest statistical source of heterogeneity (p=0.18). The limitations of biopsy as a reference standard are described above, but limiting MRI assessment to patients who

have had a prostatectomy introduces patient selection bias, as radical prostatectomy patients tend to be younger,fitter, and tend to have clinically significant tumor, prompting surgery. Radical prostatectomy allows examination of the entire gland including the anterior gland (which TRUS cannot) and detects multifocality, which is frequent (44,45). Eight of nine studies in the prostatectomy subgroup assessed for multifocal disease or for tumor in multiple segments of the prostate and this subgroup's results may be more representative of the diagnostic accuracy of high b-value diffusion.

There are limitations to this study. Limitation by language and database may have introduced bias. The use of two larger databases, and grey literature, should encompass most eligible English language studies. Publication bias was not assessed as a Deek's funnel plot is less accurate in meta-analyses with small numbers of studies (46). Finally, this study was restricted to testing localization of prostate cancer within the gland. This is important in determining the accuracy of high b-value diffusion, but not the only useful outcome. Identifying the presence of capsular breach, seminal vesicle invasion and pelvic lymphadenopathy are important staging and prognostic characteristics not assessed in this study.

In conclusion, these findings should be considered cautiously given the degree of statistical heterogeneity. However, this meta-analysis demonstrated that high b-value

diffusion is a valuable diagnostic tool, with a sensitivity of 59%, specificity of 92% and sROC AUC of 0.92. There was better diagnostic performance by visual assessment of high b-value DWI studies compared to ADC quantification.

References

1. Jemal A, Siegel R, Ward E, et al. Cancer Statistics. CA Cancer J Clin 2009;59:225–249.

2. American Cancer Society. Cancer Facts & Figures 2012. American Cancer Society. 2012.

3. Barentsz JO, Richenberg J, Clements R, et al. ESUR prostate MR guidelines 2012. Eur Radiol 2012;22:746–757.

4. Kirkham A, Haslam P, Keanie JY, et al. Prostate MRI: who, when, and how? Report from a UK consensus meeting. Clin Radiol 2013;68:1016–1023.

5. Tamada T, Sone T, Jo Y, et al. Apparent diffusion coefficient values in peripheral and transition zones of the prostate: comparison between normal and malignant prostatic tissues and correlation with histologic grade. J Magn Reson Imaging 2008;28:720–726.

6. Hambrock T, Somford DM, Huisman HJ, et al. Relationship between apparent diffusion coefficients at 3.0-T mr imaging and gleason grade in peripheral zone prostate cancer. Radiology 2011;259:453–461.

7. Rosenkrantz AB, Sigmund EE, Johnson G, et al. Prostate cancer: Feasibility and preliminary experience of a diffusional kurtosis model for detection and assessment of aggressiveness of peripheral zone cancer. Radiology 2012;264:126–135.

8. Riches SF, Hawtin K, Charles-Edwards EM, et al. Diffusion-weighted imaging of the prostate and rectal wall: Comparison of biexponential and monoexponential modelled diffusion and associated perfusion coefficients. NMR Biomed 2009;22:318–325.

9. Mazaheri Y, Afaq A, Rowe DB, et al. Diffusion-weighted magnetic resonance imaging of the prostate: Improved robustness with stretched exponential modeling. J Comput Assist Tomogr 2012;36:695–703.

10. Panagiotaki E, Chan RW, Dikaios N, et al. Microstructural characterization of normal and malignant human prostate tissue with vascular, extracellular, and restricted diffusion for cytometry in tumours magnetic resonance imaging. Invest Radiol 2015;50:218–227.

11. Kim CK, Park BK, Kim B. Diffusion-weighted MRI at 3 T for the evaluation of prostate cancer. Am J Roentgenol 2010;194:1461–1469.

12. Wang X, Qian Y, Liu B, et al. High-b-value diffusion-weighted MRI for the

detection of prostate cancer at 3 T. Clin Radiol 2014;69:1165–1170.

13. Ueno Y, Kitajima K, Sugimura K, et al. Ultra-high b-value diffusion-weighted MRI for the detection of prostate cancer with 3-T MRI. J Magn Reson Imaging 2013;38:154–160.

14. Tamada T, Kanomata N, Sone T, Jo Y, et al. High b value (2,000 s/mm2) diffusion-weighted magnetic resonance imaging in prostate cancer at 3 Tesla: comparison with 1,000 s/mm2 for tumor conspicuity and discrimination of aggressiveness. PLoS One 2014;9:e96619.

15. Rosenkrantz AB, Hindman N, Lim RP, et al. Diffusion-weighted imaging of the prostate: Comparison of b1000 and b2000 image sets for index lesion detection. J Magn Reson Imaging 2013;38:694–700.

16. Kim CK, Park BK, Kim B. High-b-value diffusion-weighted imaging at 3 T to detect prostate cancer: comparisons between b values of 1,000 and 2,000 s/mm2. AJR Am J Roentgenol 2010;194:W33-37.

17. Koo JH, Kim CK, Choi D, et al. Diffusion-weighted magnetic resonance imaging for the evaluation of prostate cancer: Optimal B value at 3T. Korean J Radiol 2013;14:61–69.

18. Wu L-M, Xu J-R, Ye Y-Q, et al. The clinical value of diffusion-weighted

imaging in combination with T2-weighted imaging in diagnosing prostate carcinoma: a systematic review and meta-analysis. AJR Am J Roentgenol 2012;199:103–110.

19. Jie C, Rongbo L, Ping T, et al. The value of diffusion-weighted imaging in the detection of prostate cancer: a meta-analysis. Eur Radiol 2014;24:1929–1941.

20. Jin G, Su DK, Luo N Bin, et al. Meta-analysis of diffusion-weighted magnetic resonance imaging in detecting prostate cancer. J Comput Assist Tomogr 2013;37:195–202.

21. Moher D, Shamseer L, Clarke M, et al. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4:1.

22. Whiting PF, Rutjes AWS, Westwood ME, et al. Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529–536.

23. Chen M, Dang H-D, Wang J-Y, et al. Prostate cancer detection: comparison of T2-weighted imaging, diffusion-weighted imaging, proton magnetic resonance spectroscopic imaging, and the three techniques combined. Acta radiol 2008;49:602–610.

24. Girometti R, Bazzocchi M, Como G, et al. Negative predictive value for cancer in patients with “gray-Zone” PSA level and prior negative biopsy: Preliminary results with multiparametric 3.0 tesla MR. J Magn Reson Imaging 2012;36:943–950.

25. Isebaert S, Van den Bergh L, Haustermans K, et al. Multiparametric MRI for prostate cancer localization in correlation to whole-mount histopathology. J Magn Reson Imaging 2013;37:1392–1401.

26. Kumar V, Jagannathan NR, Kumar R, et al. Apparent diffusion coefficient of the prostate in men prior to biopsy: determination of a cut-off value to predict malignancy of the peripheral zone. NMR Biomed 2007;20:505–511.

27. Rosenkrantz ABB, Khalef V, Xu W, et al. Does normalisation improve the diagnostic performance of apparent diffusion coefficient values for prostate cancer assessment? A blinded independent-observer evaluation. Clin Radiol 2015;70:1032–1037.

28. Vilanova JC, Barceló-Vidal C, Comet J, et al. Usefulness of prebiopsy multifunctional and morphologic MRI combined with free-to-total prostate-specific antigen ratio in the detection of prostate cancer. Am J Roentgenol 2011;196:715–722.

29. Peng Y, Jiang Y, Yang C, et al. Quantitative analysis of multiparametric prostate MR images: differentiation between prostate cancer and normal tissue and correlation with Gleason score--a computer-aided diagnosis development study. Radiology 2013;267:787–796.

30. Lim HK, Kim JK, Kim K et al. Prostate cancer: Apparent diffusion coefficient map with T2-weighted images for detection - A multireader study. Radiology 2009;250:145–151.

31. Jones CM, Athanasiou T. Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. Ann Thorac Surg 2005;79:16–20.

32. Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. Eur Urol 2016;69:16–40.

33. Ueno Y, Kitajima K, Sugimura K, et al. Ultra-high b-value diffusion-weighted MRI for the detection of prostate cancer with 3-T MRI. J Magn Reson Imaging 2013;38:154–160.

34. Kitajima K, Takahashi S, Ueno Y, et al. 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 prostat. Eur J Radiol 2013;82:1219–1226.

35. Ohgiya Y, Suyama J, Seino N, et al. Diagnostic accuracy of ultra-high-b-value 3.0-T diffusion-weighted MR imaging for detection of prostate cancer. Clin Imaging 2012;36:526–531.

36. Catalona WJ, Partin AW, Slawin KM, et al. Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: a prospective multicenter clinical trial. JAMA 1998;279:1542–1547.

37. Ahmed HU, Kirkham A, Arya M, et al. Is it time to consider a role for MRI before prostate biopsy? Nat Rev Clin Oncol 2009;6:197–206.

38. Dietrich O, Heiland S, Sartor K. Noise correction for the exact determination of apparent diffusion coefficients at low SNR. Magn Reson Med 2001;45:448–453.

39. Damon BM. Effects of image noise in muscle diffusion tensor (DT)-MRI assessed using numerical simulations. Magn Reson Med 2008;60:934–944.

40. Sinha S, Sinha U. In vivo diffusion tensor imaging of the human prostate. Magn Reson Med 2004;52:530–537.

41. Jambor I, Merisaari H, Taimen P, et al. Evaluation of different mathematical

models for diffusion-weighted imaging of normal prostate and prostate cancer using high b-values: A repeatability study. Magn Reson Med 2015;73:1988–1998.

42. Toivonen J, Merisaari H, Pesola M, et al. Mathematical models for diffusion-weighted imaging of prostate cancer using b values up to 2000 s/mm2: Correlation with Gleason score and repeatability of region of interest analysis. Magn Reson Med 2015;74:1116–1124.

43. Hall MG, Bongers A, Sved P, et al. Assessment of non-Gaussian diffusion with singly and doubly stretched biexponential models of diffusion-weighted MRI (DWI) signal attenuation in prostate tissue. NMR Biomed 2015;28:486–495.

44. Le JD, Tan N, Shkolyar E, et al. Multifocality and Prostate Cancer Detection by Multiparametric Magnetic Resonance Imaging: Correlation with Whole-mount Histopathology. Eur Urol 2014;67:569–576.

45. Noguchi M, Stamey TA, McNeal JE, et al. Prognostic Factors for Multifocal Prostate Cancer in Radical Prostatectomy Specimens: Lack of Significance of Secondary Cancers. J Urol 2003;170:459–463.

46. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. 2011.

Table 1. Principle characteristics of included studies

Study	Year	Country	No. of patients	Age (range)	PSA (range)	Design
Chen	2008	China	42	63(45–82)	52.5(4.7–147)	Retro
Girometti	2012	Italy	26	64*(51–74)	6.0*(2.5–10)	Pro
Isebaert	2013	Belgium	75	66*(49–74)	10.4(1.5–70.9)	Pro
Kim	2000	South Korea	48	66(45–80)	7.2*(2.3–23.2)	Retro
Koo	2013	South Korea	80	66(45–81)	7.2(1.2–57)	Retro
Kumar	2007	India	23	64.5	11*(0.5–1000)	Pro
Lim	2009	South Korea	52	65(48–76)	66(45–80)	Retro
Peng	2013	USA	48	61.5(44–73)	15.6(0.8–256)	Retro
Rosenkrantz	2015	USA	58	63	8.2	Retro
Vilanova	2011	Spain	70	63.5(43–87)	7.4*(4–17.2)	Retro

PSA – Prostate specific antigen (ng/mL), * Median, Pro – prospective, Retro – retrospective

Table 2. Imaging and methodological characteristics of the included studies.

Study	Field Strength	Coil	b-values (s/mm^2)	TR/TE (ms)	NSA	Voxel Size (mm)	RS	Blind	Spasm	Threshold Used
Chen	1.5T	A	0 & 1000	3200/94	-	1.8 x 1.8 x -	Bx	Y	U	ADC
Girometti	3T	A	0, 800 & 1200	4600/60	6	2.3 × 2.5 × 3.0	Bx	U	Y	ADC
Isebaert	1.5T	A	0, 50, 100, 500, 750 & 1000	4000/79	-	3.0 × 3.0 × 4.0	RP	Y	U	Visual
Kim	3T	A	0 & 1000; 0 & 2000	2,740–2,950/ 83–95	3	1.8 × 1.8 × 3.0	RP	Y	Y	Visual
Koo	3T	A	0 & 300; 0 & 700; 0 & 1000; 0 & 2000	4830-4840/ 75-76	3	1.8 × 1.8 × 3.0	RP	Y	Y	Visual Scale
Kumar	1.5T	B	0, 250, 500, 750 & 1000	3000/96	-	1.8 × 1.8 × 4.0–5.0	Bx	Y	U	ADC
Lim	3T	A	0 & 1000	3400/117	-	0.9 × 0.9 × 4.0	RP	Y	Y	Visual scale
Peng	1.5T	B	0, 50, 200, 1500 & 2000; 0 & 1000*	2948–8616/ 71–85	1-4	0.81 × 0.81-1.28 × 1.28	RP	U	Y	ADC
Rosenkrantz	3T	A	50 & 1000	4100/86	10	2.0 × 0.9 × 3.0	RP	Y	U	ADC
Vilanova	1.5T	B	0 & 1000	8250/94	6	2.0 × 2.0 × 3.0	C	Y	U	ADC

Coil A – without endorectal coil, *Coil B* – with endorectal coil, *TR* – repetition time, *TE* – Echo time, *NSA* – Number of signal averages, *RS* – reference standard, *RP* – radical prostatectomy, *Bx* – biopsy, *C* – both RP and Bx included, *Blind* – Blinded, Y – yes, U – unclear, *Spasm* – Anti-spasmodics, *Th* – Threshold, *ADC* – Apparent diffusion coefficient

* 29 patients were imaged with b-values of 0,50, 200, 1500 and 2000; 24 patients were imaged with b-values of 0 and 1000.

Table 3. Results of the subgroup analysis.

Study characteristics	No	Pooled sensitivity (95% CI)	Pooled specificity (95% CI)	AUC	p-value*
Total	13	0.59 (0.57–0.91)	0.92 (0.91–0.92)	0.92	
b-value					0.31
1000	9	0.55 (0.53–0.58)	0.90 (0.89–0.92)	0.91	
>1000	3	0.72 (0.67–0.76)	0.94 (0.93–0.95)	0.98	
Field strength					0.20
1.5T	6	0.49 (0.46–0.51)	0.90 (0.88–0.91)	0.89	
3T	7	0.74 (0.71–0.79)	0.93 (0.92–0.94)	0.96	
Coil					0.32
With endorectal	4	0.76 (0.71–0.80)	0.86 (0.83–0.89)	0.84	
Without endorectal	9	0.56 (0.53–0.58)	0.93 (0.92–0.93)	0.94	
Reference standard					0.16
Biopsy	3	0.73 (0.63–0.82)	0.86 (0.82–0.89)	0.86	
Prostatectomy	9	0.56 (0.56–0.60)	0.92 (0.91–0.93)	0.94	
Threshold method					0.03
ADC	7	0.64 (0.59–0.69)	0.87 (0.84–0.89)	0.88	
Visual	6	0.58 (0.55–0.60)	0.93 (0.92–0.93)	0.95	
Patient selection bias					0.29
High risk	10	0.58 (0.56–0.60)	0.92 (0.91–0.93)	0.94	
Low risk	3	0.77 (0.68–0.84)	0.83 (0.79–0.87)	0.86	

AUC – Area under the curve. *p-value – comparison of diagnostic odds ratio of subgroups

Figures

Fig. 1. QUADAS-2 results summarizing the proportion of low, high and unclear risk of bias and applicability concerns.

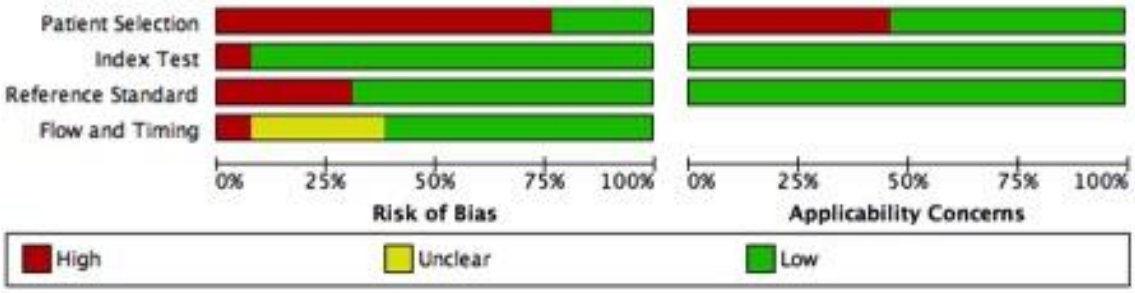


Fig. 2. Forest plot of sensitivity with pooled sensitivity, Q statistic of the chi-squared, and I-squared results.

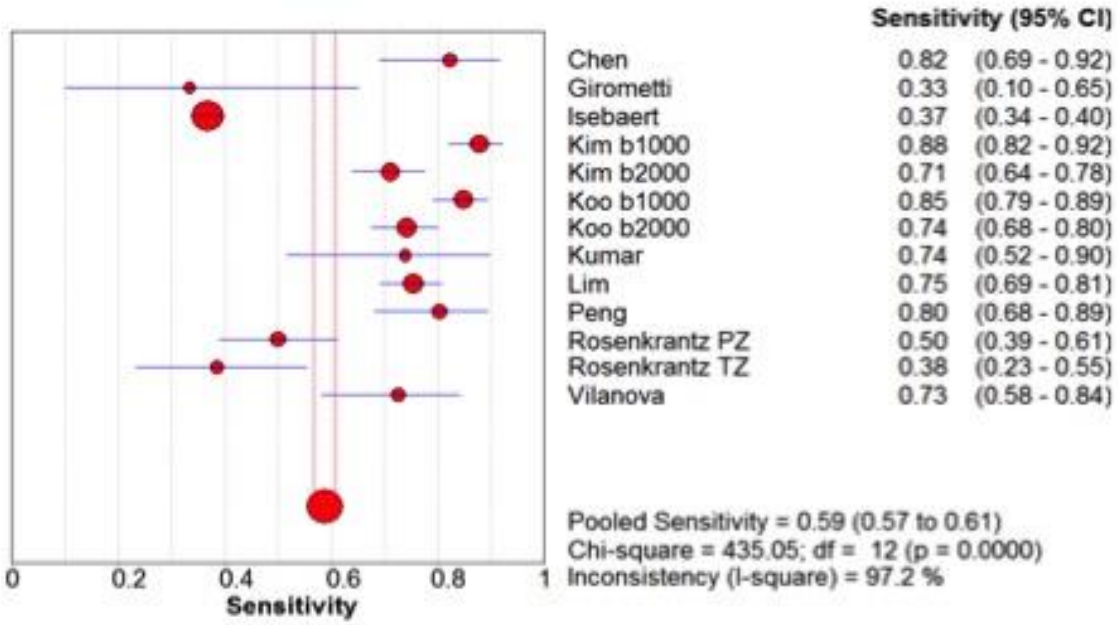


Fig. 3. Forest plot of specificity with pooled specificity, Q-statistic of the chi-squared, and I-squared results.

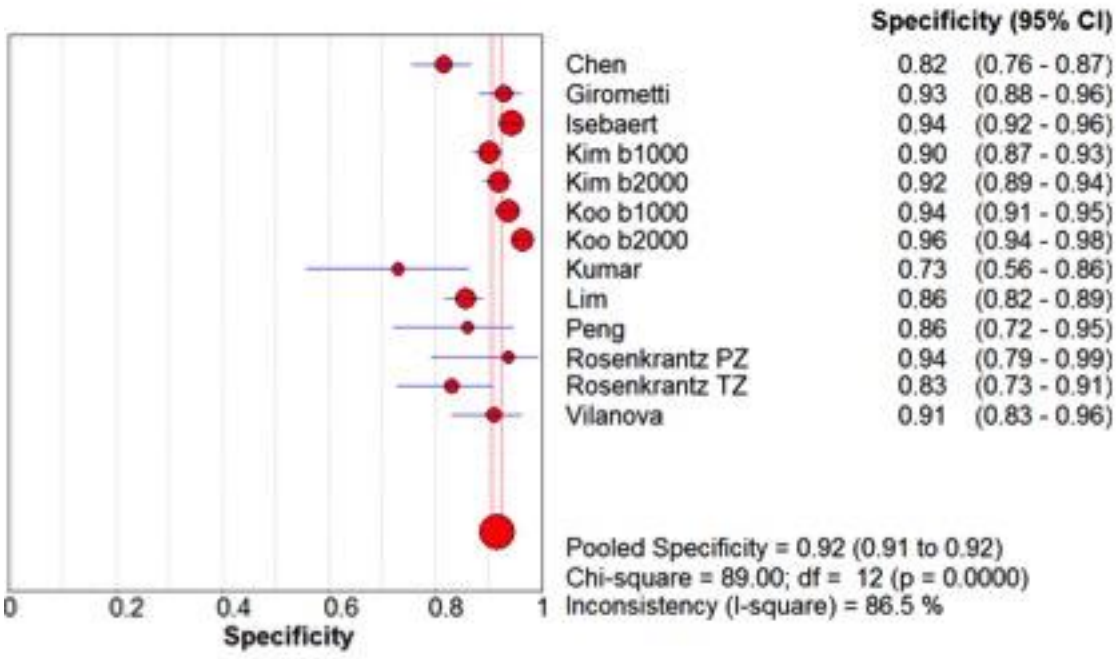
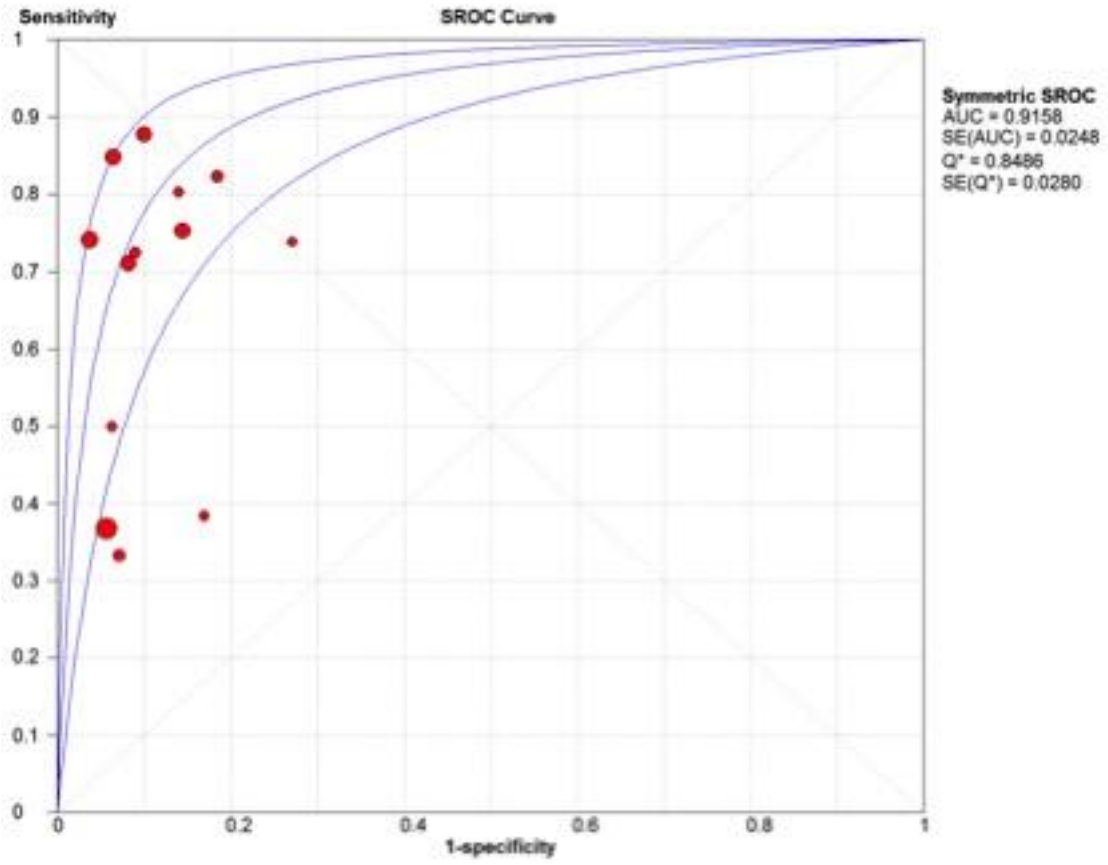


Fig. 4. The summary Receiver Operating Characteristic (sROC) curve for high b-value DWI in detecting prostate cancer.


Sensitivity
SROC Curve
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
1-specificity
Symmetric SROC
AUC = 0.9158
SE(AUC) = 0.0248
Q* = 0.8486
SE(Q*) = 0.0280

Supplementary Table 1. MEDLINE search strategy

	Search	Number of studies
1	"prostatic neoplasms/diagnosis"[MeSH Terms]	54849
2	“diffusion magnetic resonance imaging” [MeSH Terms]	13053
3	1 AND 2	402
4	Limit 3 to Humans	398
5	Limit 4 to English	372
6	Limit 5 to Abstracts	351

Supplementary Table 2. Quality assessment of the included studies.

	Risk of bias				Applicability concerns		
	Patient Selection	Index test(s)	Reference Test	Flow & Timing	Patient Selection	Index test(s)	Reference Test
Chen	+	+	-	+	+	+	+
Girometti	-	+	-	-	-	+	+
Isebaert	-	+	+	U	+	+	+
Kim	-	+	+	+	-	+	+
Koo	-	+	+	+	-	+	+
Kumar	+	+	-	+	+	+	+
Lim	-	+	+	+	+	+	+
Peng	-	-	+	U	-	+	+
Rosenkrantz	-	+	+	U	+	+	+
Vilanova	+	+	-	+	+	+	+

+ Low risk, – High risk, U Unclear risk

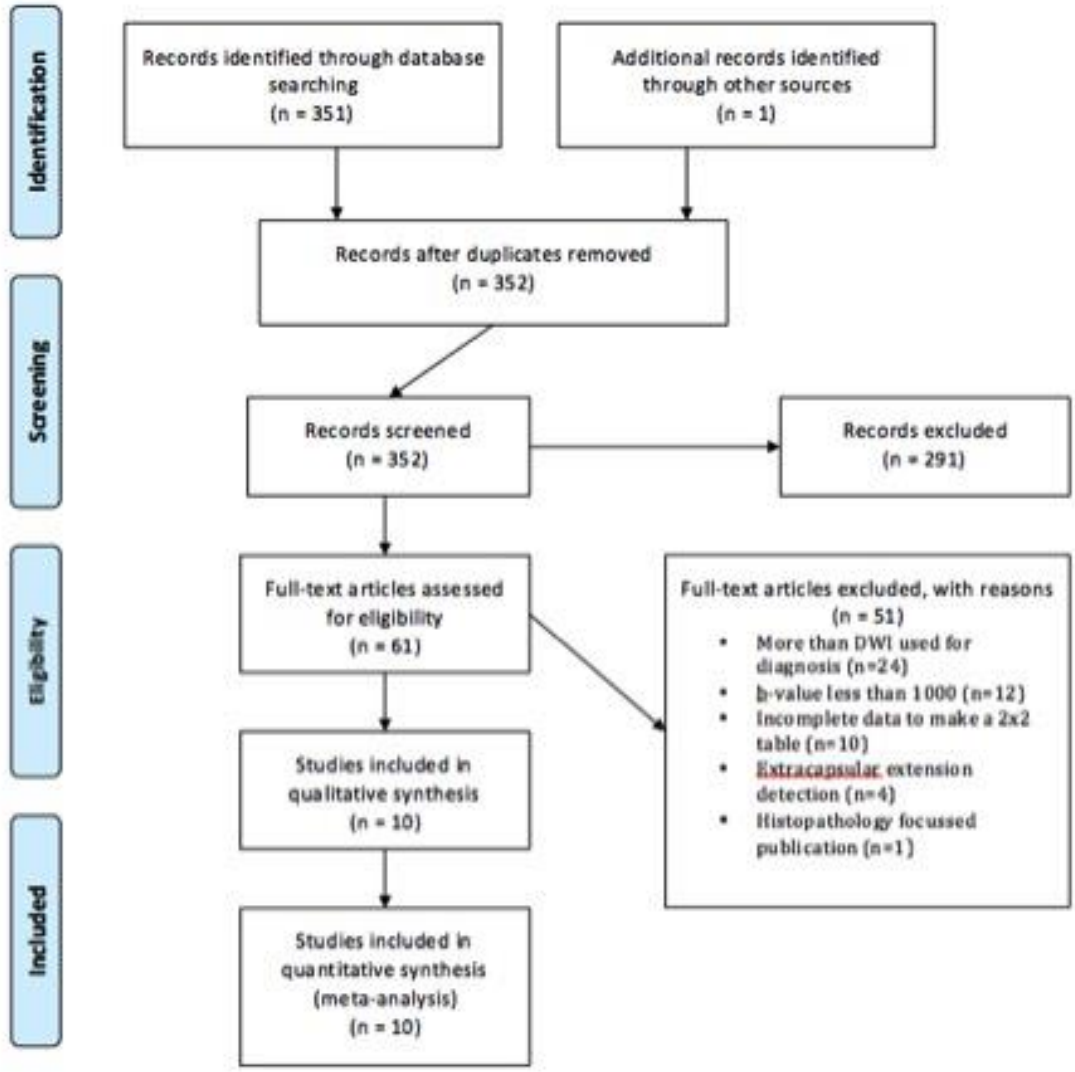
Supplementary Table 3. Diagnostic performance of eligible studies and subsets.

Study	TP	FP	FN	TN	Sens	Spec	Notes
Chen	42	37	9	164	0.82	0.82	
Girometti	4	14	8	182	0.33	0.93	
Isebaert	359	44	617	732	0.37	0.94	
Kim	158	49	22	443	0.88	0.92	b=1000
	128	40	52	452	0.71	0.92	b=2000
Koo	174	38	31	557	0.85	0.94	b=1000
	152	22	53	573	0.74	0.96	b=2000
Kumar	17	10	6	27	0.63	0.82	
Lim	171	57	56	340	0.75	0.86	
Peng	49	6	12	37	0.86	0.80	
Rosenkrantz	15	13	24	64	0.39	0.83	TZ
	42	2	42	30	0.50	0.94	PZ
Vilanova	37	8	14	81	0.73	0.91	

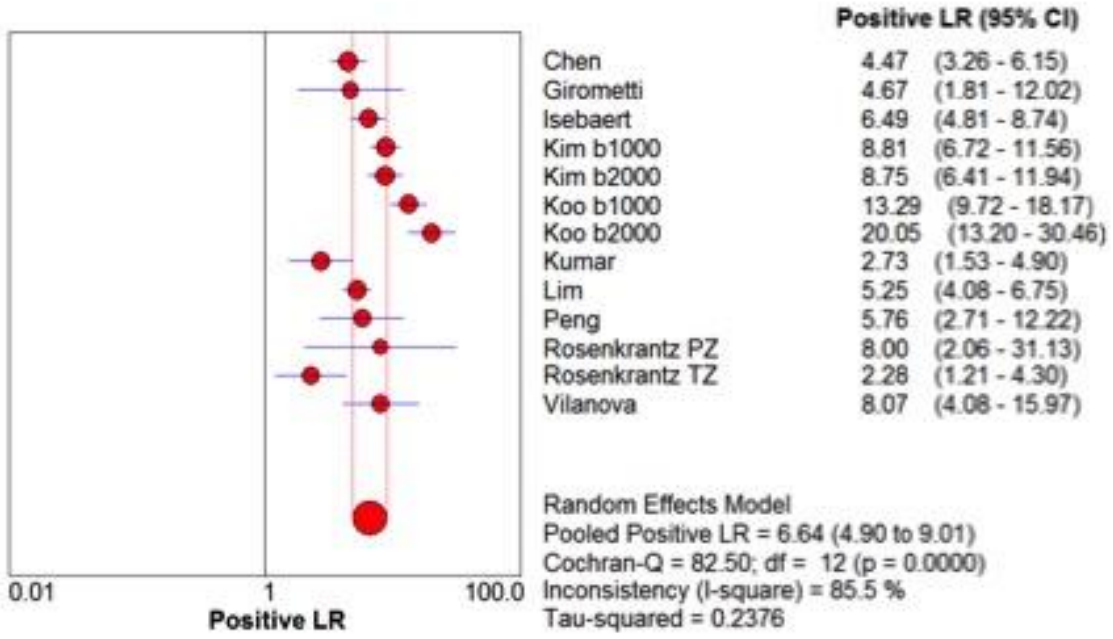
TP – true positives, FP – false positives, FN – false negatives, TN – true negatives, Sens – sensitivity, Spec – specificity, TZ – transitional zone, PZ – peripheral zone

Supplementary Figure Legends

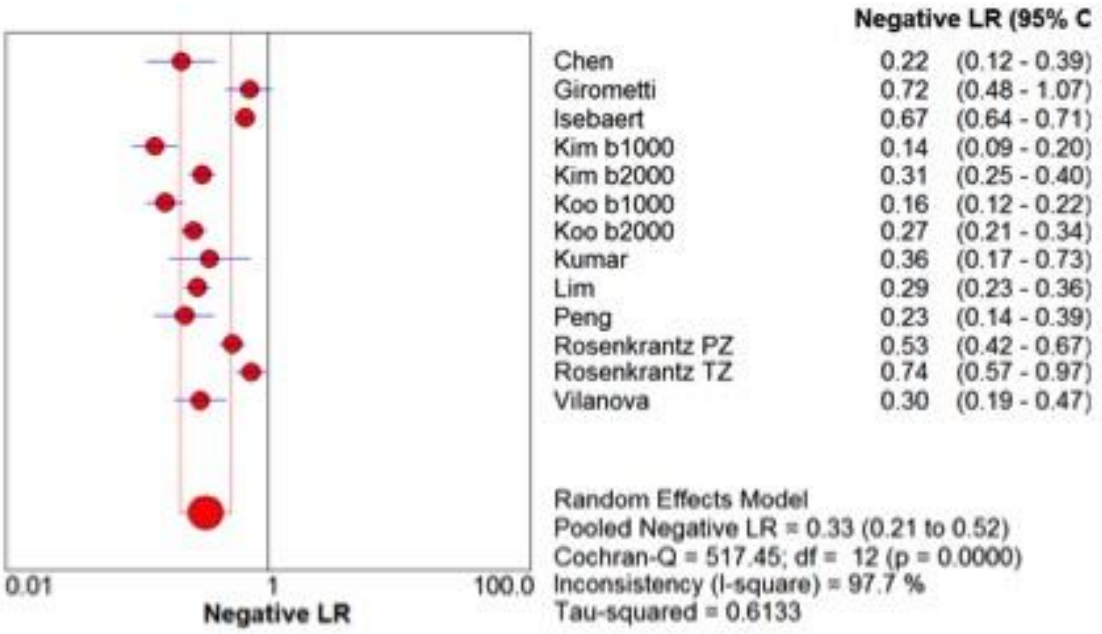
Supplementary Fig. 1. Flow diagram for articles identified and included in this meta-analysis.



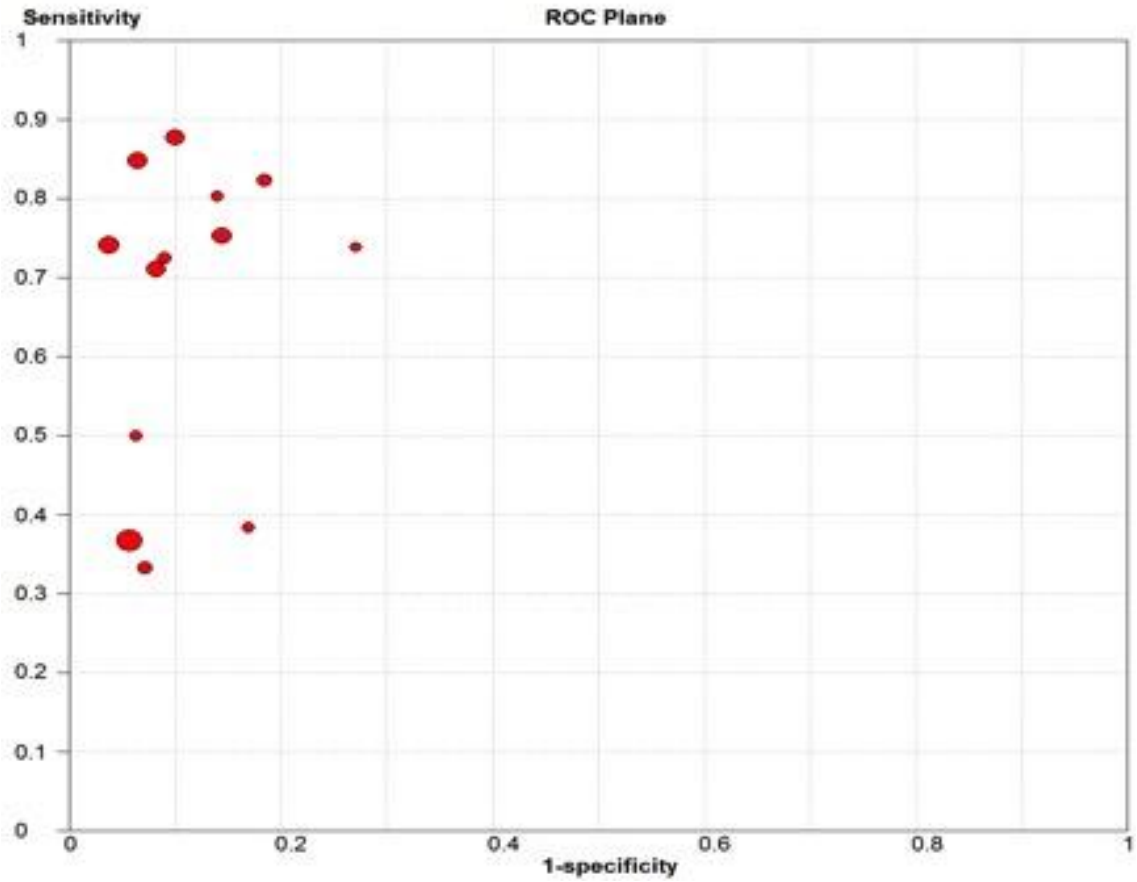
Supplementary Fig. 2. Forest plot of positive likelihood ratio with pooled positive likelihood ratio, Q statistic of the chi-squared, and I-squared results.



Supplementary Fig. 3. Forest plot of negative likelihood ratio with pooled negative likelihood ratio, Q statistic of the chi-squared, and I-squared results.



Supplementary Fig. 4. Sensitivity and 1-specificity plotted in Receiver Operating Characteristic curve for individual studies and subsets.


Sensitivity
ROC Plane
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
1-specificity