

BMJ Open Diffusion-weighted MRI in differentiating malignant from benign thyroid nodules: a meta-analysis

Lihua Chen,^{1,2} Jian Xu,³ Jing Bao,⁴ Xuequan Huang,¹ Xiaofei Hu,¹ Yunbao Xia,² Jian Wang¹

To cite: Chen L, Xu J, Bao J, *et al.* Diffusion-weighted MRI in differentiating malignant from benign thyroid nodules: a meta-analysis. *BMJ Open* 2016;**6**:e008413. doi:10.1136/bmjopen-2015-008413

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2015-008413>).

LC and JX are co-first authors.

Received 7 April 2015
Revised 2 September 2015
Accepted 23 October 2015



¹Department of Radiology, Southwest Hospital, Third Military Medical University, Chongqing, China

²Department of Radiology, Taihu Hospital, Wuxi, China

³Department of General Surgery, Taihu Hospital, Wuxi, China

⁴Molecular Biology Lab, Wuxi Center for Disease Control and Prevention, Wuxi, China

Correspondence to

Dr Jian Wang;
wangjian811@gmail.com
or
Yunbao Xia;
xiayb@aliyun.com;

ABSTRACT

Objectives: To perform a meta-analysis to evaluate the diagnostic efficacy of diffusion-weighted imaging (DWI) in differentiating malignant from benign thyroid nodules.

Design: A meta-analysis.

Data sources and study selection: Medical and scientific literature databases were searched for original articles published up to August 2015. Studies were selected if they (1) included diagnostic DWI for differentiating malignant from benign thyroid lesions, (2) included patients who later underwent biopsy and (3) presented sufficient data to enable the construction of contingency tables.

Data synthesis: For each study, the true-positive, false-positive, true-negative and false-negative values were extracted or derived, and 2×2 contingency tables were constructed. Methodological quality was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) instrument. The heterogeneity test, threshold effect test, subgroup analyses and publication bias analyses were performed.

Results: From the 113 identified search results, 15 studies, representing a total of 765 lesions, were included in the meta-analysis. We detected heterogeneity between studies but found no evidence of publication bias. The pooled weighted sensitivity was 0.90 (95% CI 0.85 to 0.93); the specificity was 0.95 (95% CI 0.88 to 0.98); the positive likelihood ratio was 16.49 (95% CI 7.37 to 36.86); the negative likelihood ratio was 0.11 (95% CI 0.08 to 0.16); and the diagnostic OR was 150.73 (95% CI 64.96 to 349.75). The area under the receiver operator characteristic curve was 0.95 (95% CI 0.93 to 0.97).

Conclusions: Quantitative DWI may be a non-invasive, non-radiative and accurate method of distinguishing malignant from benign thyroid nodules. Nevertheless, large-scale trials are necessary to assess its clinical value and to establish standards regarding b values and cut-off values for DWI-based diagnosis.

INTRODUCTION

Thyroid nodules, the most common pathology involving the thyroid gland, consist of

Strengths and limitations of this study

- The Preferred Reporting Items for Systematic reviews and Meta-Analyses statement was used to improve the reporting of our research.
- Hierarchical summary receiver operating characteristic (HSROC) curves were constructed to assess sensitivity (SEN) and specificity (SPE).
- We presented a new point: using a high b value may provide higher diagnostic accuracy.
- Studies included in our meta-analysis lacked a description of apparent diffusion coefficient reproducibility and the sample size of included studies was relatively small.

discrete lesions within the thyroid gland that are often palpable and typically sonographically distinct from the surrounding thyroid parenchyma.¹ Less than 5% of palpable thyroid nodules are malignant; however, these nodules must be distinguished from benign thyroid nodules to correctly and efficaciously treat patients suffering from this pathology.²

Because clinical findings do not provide a definitive diagnosis, several useful, non-invasive imaging tests (such as ultrasonography (US) and radionuclide scintigraphy) can be used to determine which nodules should be histopathologically evaluated to rule out the possibility of thyroid malignancy. US has been used as a first step in the assessment of these nodules, but no single US criterion has been demonstrated to accurately differentiate benign nodules from malignant nodules. Furthermore, the hazards associated with radiation exposure during radionuclide scintigraphy are unavoidable, and some functioning nodules (hot nodules) found on scintigraphy are malignant.

Diffusion-weighted imaging (DWI) is a type of functional MRI that is based on the diffusion of water molecules through the tissue of interest (ie, tumour tissue). DWI

can provide crucial information regarding the molecular profile of the underlying pathology and pathophysiological mechanisms.³ Specifically, the diffusion of water molecules in malignant tumours is restricted, which results in a decreased apparent diffusion coefficient (ADC); this difference in the ADC facilitates the differentiation of benign tumours from malignant tumours.⁴

Many studies^{1 5–18} have shown that DWI has the potential to differentiate benign from malignant thyroid nodules. However, the sample sizes of these studies were relatively small, and the findings have been inconclusive. The aim of this study was to systematically review all of the studies related to the ability of DWI to differentiate benign from malignant thyroid nodules. Moreover, based on the extracted data, an analysis of the technical aspects of DWI and its additional value for tumour characterisation is presented.

METHODS

As this meta-analysis was based on previously published studies, ethical approval was not necessary.

Literature search

The PubMed, EMBASE, Cochrane Library and China National Knowledge Infrastructure (CNKI) databases were searched by two independent observers. The terms 'Diffusion-Weighted Imaging (MeSH)' or 'DWI' were used for the diagnostic test, and the terms 'thyroid nodules (MeSH)', 'thyroid lesions', or 'thyroid' were used for the clinical domain (Search Strategy S1). We limited our search to publications that met the following criteria: published in the English or Chinese language; the presence of the search term within the title or abstract of the article; and a publication date no later than May 2014. Review articles, letters, comments, case reports and unpublished articles were excluded. The reference lists of all retrieved articles were manually cross-checked.

Selection of articles

Two authors initially screened the titles and abstracts of the search results and retrieved the full texts of all potentially relevant reports. Next, the authors independently reviewed all relevant reports according to the pre-defined inclusion criteria. Disagreements were resolved by consensus or arbitration by a third author, who assessed all of the involved items. The majority opinion was used to determine whether a particular study met the selection criteria.

Studies were considered as eligible if the following criteria were met: (1) the study included a diagnostic DWI for differentiating malignant from benign thyroid lesions; (2) the controls underwent histopathological analysis (surgery/biopsy) and/or follow-up analysis and (3) the data were sufficient to accurately determine the true-positive or false-negative results.

Studies were excluded if (1) there were fewer than 20 patients; (2) multiple reports were published for the same study population (in this case, the most detailed or recent publication was chosen) or (3) the study included patients who had previously undergone treatment for thyroid lesions.

Quality assessment and data extraction

The aforementioned three authors extracted data from the selected reports. The methodological quality of the included studies was independently assessed by two observers using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool, which was specifically developed to systematically review the diagnostic accuracy of studies.^{19–21} Additionally, relevant data, including author, the study nation, the description of the study population, the study design characteristics, the magnetic field strength, the pulse sequences and descriptions of the interpretations of the diagnostic tests, were extracted from each study. To resolve disagreements between the reviewers, a third reviewer assessed the disputed material, and the majority opinion was used in the analysis.

For each study, the estimated true-positive (TP), false-positive (FP), true-negative (TN) and false-negative (FN) values, sensitivity (SEN), specificity (SPE), positive likelihood ratio (PLR) and negative likelihood ratio (NLR) for the detection of lesions were extracted, and 2x2 contingency tables were constructed.

Meta-analysis

Exploring study heterogeneity is important in understanding the possible factors that influence accuracy estimates and in evaluating the appropriateness of statistical pooling of accuracy estimates from various studies. Visual inspection of the forest plots, standard χ^2 -tests and the inconsistency index (I^2) were used to estimate the heterogeneity of the individual studies, using Stata software (Stata Corporation, College Station, Texas, USA). $p < 0.1$ or $I^2 > 50\%$ suggested notable heterogeneity.²² If notable heterogeneities were detected, the relevant data were pooled using a random-effects coefficient binary regression model; otherwise, a fixed-effects coefficient binary regression model was used.²³

In the diagnostic test, one primary cause of heterogeneity is the threshold effect, which arises when different cut-off values, or thresholds, are used to define a positive (or negative) test result between different studies. When a threshold effect exists, there is a negative correlation between SEN and SPE.^{24–26} The Spearman correlation coefficient between the logit of SEN and the logit of (1 – SPE) was computed to assess the threshold effect, using Meta-Disc V.1.4; a strongly positive correlation ($p < 0.05$) suggests a threshold effect. We constructed hierarchical summary receiver operating characteristic (HSROC) curves to assess SEN and SPE.²⁷ The areas under the ROC curves (AUCs) were used to analyse the

diagnostic precision of DWI in differentiating thyroid nodules.

In addition to the threshold effect on systematic review results, several other factors can result in variations in accuracy estimates between different test accuracy studies. In this study, meta-regression was used to identify such heterogeneity by comparing the accuracy measurement to study-level covariates (study nation, study design, MRI field strength, reference standard, enrolment, disease spectrum, patient spectrum or b value). Then, subgroup analyses were performed. Stratification was performed according to the following parameters: (1) the b value; (2) studies with a prospective or retrospective design; (3) magnetic field strength; (4) reference standard and (5) enrolment.

Using Stata software, the presence of publication bias was assessed by producing a Deeks funnel plot and performing an asymmetry test. Publication bias^{28 29} was considered to be present if there was a non-zero slope coefficient ($p < 0.05$), which suggested that only small studies reporting high accuracy had been published; alternatively, $p > 0.1$ suggested that there was no evidence of notable publication bias.

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement³⁰ was used to improve the reporting of our research (figure 1 and Checklist S2).

RESULTS

The database search initially yielded 113 potential literature citations; 4 additional results were identified by searching the grey literature (figure 1). After reviewing the titles and abstracts, 48 of these studies were excluded because they were duplicate publications or not relevant to this analysis. After reading the full texts, 19 of the

remaining 34 articles were excluded because (1) the patients had previously undergone treatment, (2) the article lacked sufficient information to complete a 2×2 contingency table or (3) the study was not published in English or Chinese. Following the final screening process, 15 published studies were considered to have met all of our inclusion and exclusion criteria. The data abstracted from these individual studies are summarised in table 1. The quality was moderate in 15 studies, according to the QUADAS-2 items; the results for the distribution of the study design are shown in figure 2.

Significant heterogeneity was found based on the pooled analysis ($I^2=54.5%$, $p=0.055$). Therefore, SEN, SPE, PLR and NLR were pooled using a random-effects coefficient binary regression model. The pooled weighted values were as follows: SEN 0.90 (95% CI 0.85 to 0.93); SPE 0.95 (95% CI 0.88 to 0.98); PLR 16.49 (95% CI 7.37 to 36.86); NLR 0.11 (95% CI 0.08 to 0.16); and diagnostic OR (DOR) 150.73 (95% CI 64.96 to 349.75). The AUC was 0.95 (95% CI 0.93 to 0.97). The forest plots and HSROC curves for the 15 studies are shown in figures 3–5.

A Spearman rank correlation was performed as a further assessment of the threshold effect; the Spearman correlation coefficient was determined to be 0.081 ($p=0.775$). This result indicated that no notable threshold effect was detected in the accuracy estimates among individual studies.

The results of meta-regression indicated that study nation, study design, MRI field strength, reference standard, enrolment, disease spectrum, patient spectrum and b values were not strongly correlated with accuracy. The estimated SEN and SPE for each subgroup are presented in table 2.

The results of the Deeks funnel plot asymmetry test ($p=0.786$) showed no evidence of notable publication bias (figure 6).

DISCUSSION

Thyroid nodules are highly prevalent and clinically difficult to manage. Compared with benign thyroid nodules, malignant thyroid nodules have larger nuclei, denser stroma and higher cell counts, all of which lead to increased cellularity and reduced extracellular space.³¹ Many studies^{11–14 18 32 33} and a systematic review³⁴ demonstrated that the ADCs of malignant thyroid nodules are significantly smaller than those of benign nodules. A meta-analysis³⁵ that included seven studies regarding the potential of DWI to differentiate between malignant and benign thyroid nodules was published in 2014, and suggested that DWI can be used as a diagnostic tool to distinguish benign from malignant thyroid nodules by measuring the ADC. In our study, eight additional references were included that were not present in the aforementioned meta-analysis. Unlike the previous meta-analysis, our meta-analysis examined the technical aspects of DWI and its additional value in tumour

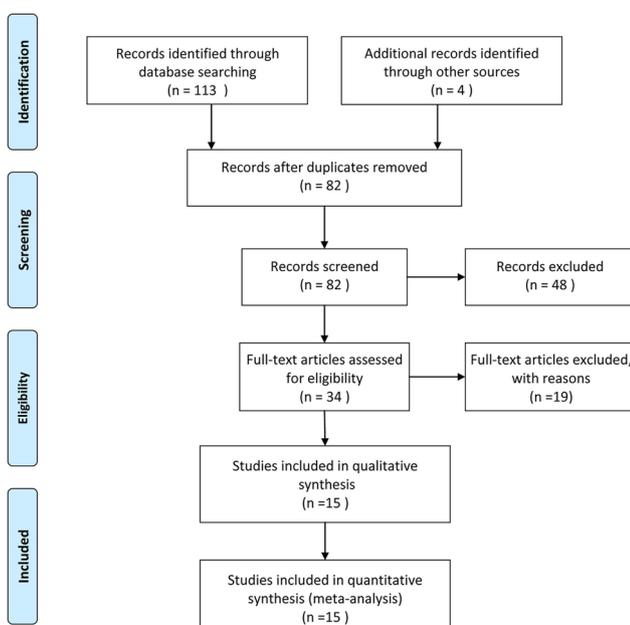


Figure 1 Flow chart illustrating the selection of studies.

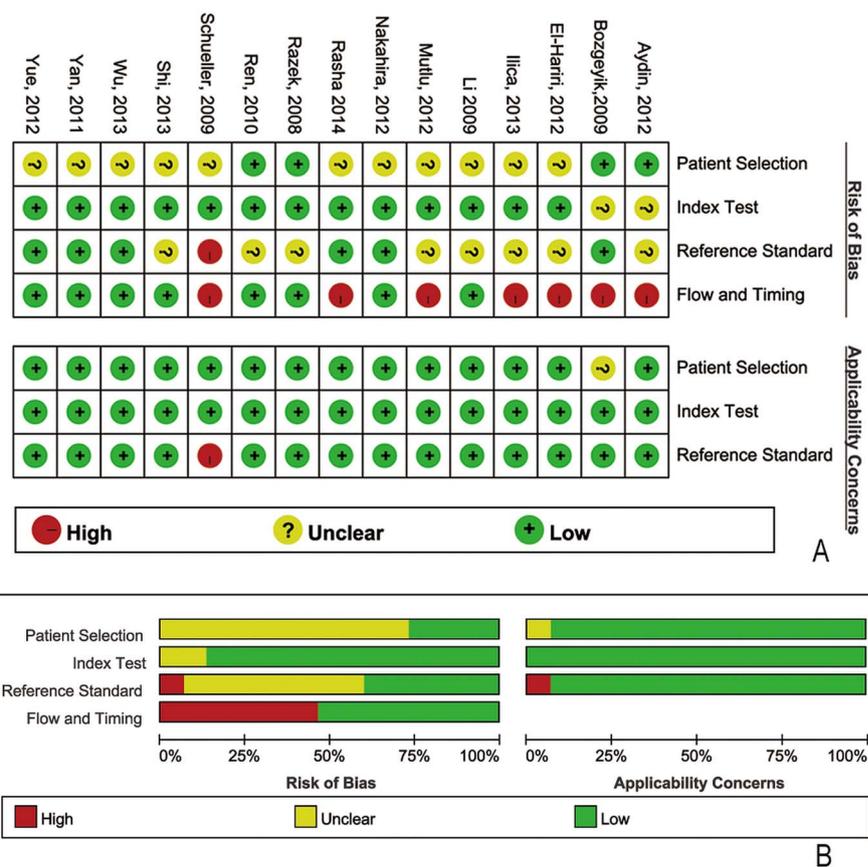
Table 1 Characteristics of the included studies

Study	Year	Nation	SEN (%)	SPE (%)	Field (T)	Enrolment	Design	Reference	Blinding	b value
Razek <i>et al</i> ¹⁸	2008	Egypt	98	92	1.5	Consecutive	pro	Histopathologic	ND	500
Li <i>et al</i> ¹⁷	2009	China	79	86	1.5	ND	retro	Histopathologic	ND	150
			86	77	1.5	ND	retro	Histopathologic	ND	300
			93	58	1.5	ND	retro	Histopathologic	ND	500
Schueller-Weidekamm <i>et al</i> ¹⁵	2009	Austria	85	100	1.0	ND	pro	Both	Unblinded	800
Bozgeyik <i>et al</i> ¹	2009	Turkey	89	100	1.5	Consecutive	pro	FNAB	Blinded	100
			100	80	1.5	Consecutive	pro	FNAB	Blinded	200
			90	100	1.5	Consecutive	pro	FNAB	Blinded	300
Ren <i>et al</i> ¹⁶	2010	China	90	90	1.5	Consecutive	retro	Histopathologic	ND	100
			83	90	1.5	Consecutive	retro	Histopathologic	ND	200
			93	83	1.5	Consecutive	retro	Histopathologic	ND	300
			93	97	1.5	Consecutive	retro	Histopathologic	ND	400
Yan <i>et al</i> ¹⁴	2011	China	87	100	1.5	ND	retro	Histopathologic	Blinded	500
Aydin <i>et al</i> ¹³	2012	Turkey	92	75	1.5	Consecutive	retro	Histopathologic	ND	400
El-Hariri <i>et al</i> ¹²	2012	Egypt	94	95	1.5	ND	pro	Both	ND	500
Mutlu <i>et al</i> ¹¹	2012	Turkey	80	97	1.5	ND	pro	Both	ND	1000
Nakahira <i>et al</i> ¹⁰	2012	Japan	94	83	1.5	ND	retro	Histopathologic	Blinded	1000
Yue <i>et al</i> ⁹	2012	China	86	79	1.5	ND	retro	Histopathologic	Blinded	300
Ilica <i>et al</i> ⁸	2013	Turkey	90	100	3	ND	pro	Both	ND	1000
Shi <i>et al</i> ⁷	2013	China	92	88	1.5	ND	retro	Histopathologic	ND	500
Wu <i>et al</i> ⁶	2013	China	77	100	1.5	ND	retro	Histopathologic	Blinded	300
			68	64	1.5	ND	retro	Histopathologic	Blinded	500
			54	71	1.5	ND	retro	Histopathologic	Blinded	800
Elshafey <i>et al</i> ⁵	2013	Egypt	96	92	1.5	ND	pro	Both	Blinded	1000

Histopathologic and FNAB.

FNAB, Fine-needle aspiration biopsy; ND, not mentioned; pro, prospective; retro, retrospective; SEN, sensitivity; SPE, specificity.

Figure 2 Methodological quality of the 15 included studies. (A) Methodological quality graph: each methodological quality item is presented as the percentages across all included studies. (B) Methodological quality summary.



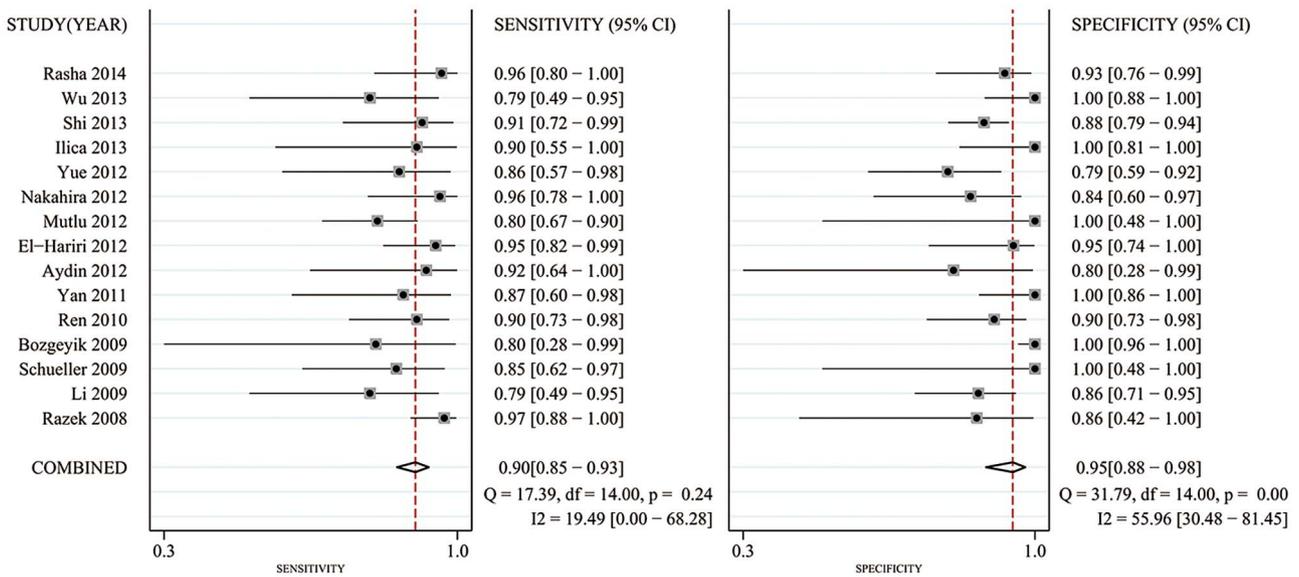


Figure 3 Forest plots of sensitivity (SEN) and specificity (SPE) with corresponding 95% CIs for diffusion-weighted image in the detection of thyroid nodules.

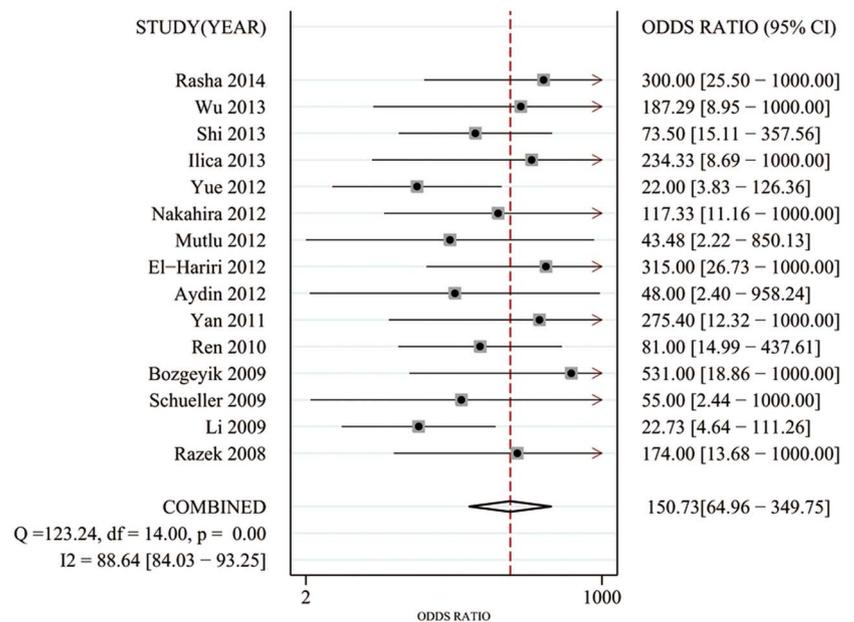
characterisation. Our results revealed for the first time that using a high b value may provide better results. The results of our meta-analysis showed that the pooled weighted SEN and SPE of the 15 included studies were 0.90 (95% CI 0.85 to 0.93) and 0.95 (95% CI 0.88 to 0.98), respectively. These results demonstrated that DWI has a high SEN and SPE for differentiating malignant from benign thyroid nodules.

The DOR represents the ratio of the odds of correctly diagnosing the diseased patients (true-positives) relative to the odds of obtaining a positive result among the non-diseased patients (false-positives). The DOR is closely linked to existing indicators and is particularly applicable to meta-analyses of diagnostic test

performance. Because the DOR is derived from logistic models, it is possible to include additional variables to correct for heterogeneity.³⁶ In our meta-analysis, we found that the estimated DOR for DWI was 150.73 (95% CI 64.96 to 349.75). This result indicated that DWI is an accurate modality for detecting malignant thyroid lesions.

We observed that the Spearman correlation coefficient was 0.081 ($p=0.775$); this result indicated that no significant threshold effect was detected. Additionally, our results indicated that none of the factors potentially impacting the meta-regression analysis results contributed to the observed heterogeneity. To determine whether there were other sources of heterogeneity, a

Figure 4 Forest plots of the diagnostic OR (DOR) with corresponding 95% CIs for diffusion-weighted image in the detection of thyroid nodules.



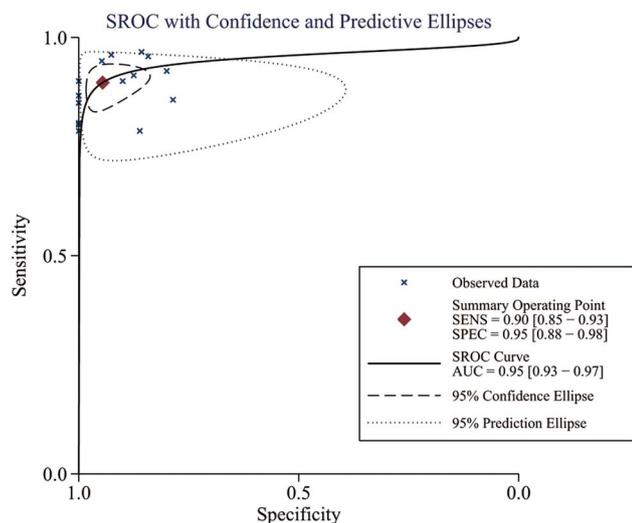


Figure 5 Hierarchical summary receiver operating characteristic (HSROC) curves from the bivariate model of diffusion-weighted image in the detection of thyroid nodules.

subgroup analysis must be performed to detect the factors that impact heterogeneity.

The b value is a very important factor affecting image quality and ADC measurements. When low b values are applied, the ADCs tend to be higher due to the contribution of perfusion. Applying high maximum b values may be preferable when ADC measurements are performed to differentiate malignant from benign tissues exclusively based on their water diffusion characteristics. However, the signal-to-noise ratio decreases as the b value increases, thus limiting the maximum b value. Six of the studies included in this review used three or more pairs of b values to compute the ADC. Three of these

six studies demonstrated that lower b values had a higher SEN and accuracy for differentiating benign from malignant nodules, whereas the remaining studies reported the opposite results. In our subgroup analysis, the results demonstrated that the pooled DORs of the 300, 500 and 1000 s/m^2 subgroups were 47.04 (95% CI 11.55 to 190.54), 53.13 (95% CI 17.95 to 219.34) and 115.21 (95% CI 28.42 to 298.76), respectively. The diagnostic accuracy may be greater in the higher b value subgroups than in the lower b value subgroups. However, there were no notable differences in the DOR or the AUC between these subgroups.

Our study contains several inherent limitations that should be considered when interpreting our results. First, most of the studies included in our meta-analysis lacked a description of ADC reproducibility and were performed in Asian countries, and no studies were from Europe or North America. Some studies³⁷⁻³⁸ have noted that selective reporting is higher among Chinese studies than elsewhere, across several fields. This issue may represent one source of heterogeneity. Second, the sample size of these studies was relatively small, which is a particular problem in diagnostic studies.³⁹ This limitation may result in an overestimation of the diagnostic accuracy, particularly in studies including non-representative samples of patients and invalid reference standards.⁴⁰ Third, our meta-analysis was based only on published studies, which are prone to report positive or significant results; the studies in which results are not significant or negative are often rejected or not even submitted. Although it is suggested that the quality of the data reported in articles accepted for publication in peer-reviewed journals is superior to the quality of unpublished data,⁴¹ including only published studies may ultimately lead to reporting bias.

Table 2 Sensitivity and specificity estimates for each subgroup

Subgroup	Number of studies	Mean SEN (%)	Mean SPE (%)	DOR	AUC (%)
b value (s/m^2)					
300	4	90 (80 to 96)	88 (83 to 93)	47 (11 to -190)	91 (88 to 95)
500	6	92 (86 to 96)	82 (77 to 86)	53 (12 to 239)	93 (90 to 96)
1000	4	88 (81 to 93)	93 (84 to 98)	115 (30 to 446)	96 (93 to 99)
Magnetic field strength (T)					
1.0	1	85	100	NA	NA
1.5	13	89 (85 to 93)	88 (85 to 91)	64 (26 to 156)	95 (91 to 99)
3.0	1	90	100	NA	NA
Study design					
Retrospective	8	91 (86 to 94)	83 (78 to 88)	33 (13 to 85)	93 (89 to 97)
Prospective	7	87 (81 to 92)	98 (94 to 99)	153 (76 to 446)	97 (95 to 99)
Blinding					
Yes	9	90 (85 to 93)	88 (83 to 92)	69 (32 to 147)	95 (93 to 97)
Unknown or no	6	88 (79 to 94)	90 (85 to 94)	66 (12 to 365)	94 (89 to 98)

The numbers in parentheses are the 95% CIs.

AUC, area under the curve; DOR, diagnostic OR; NA, not applicable; SEN, sensitivity; SPE, specificity.

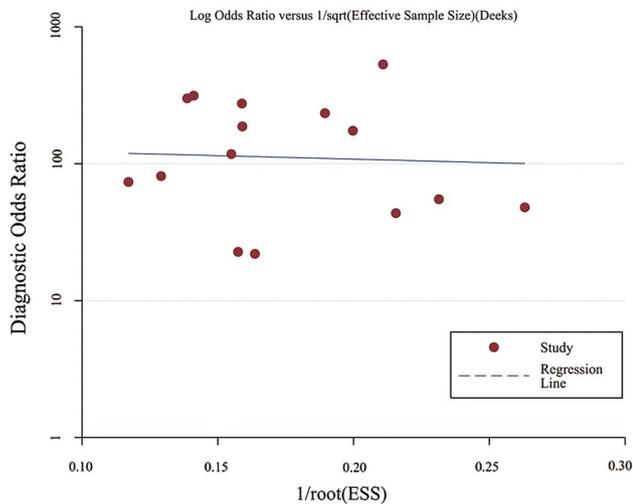


Figure 6 The funnel plot of publication bias. Linear regression of the inverse root of the effective sample size (ESS) against the log DOR (diagnostic OR) was performed to assess funnel plot asymmetry.

CONCLUSIONS

In conclusion, DWI has a high SEN and SPE, and may be a reliable, non-invasive and non-radiative imaging modality for the detection of thyroid nodules. Using a high b value may provide higher diagnostic accuracy. Nevertheless, large-scale trials are necessary to assess the clinical value of DWI, and to establish standards regarding b values and cut-off values for DWI-based diagnosis.

Acknowledgements The authors thank AJE for help in English language editing.

Contributors LC, JX, YX and JW conceived and designed the experiments. LC, JX and JW performed the experiments. JB, XH and XH analysed the data. JB, XH, XH and YX contributed reagents/materials/analysis tools. LC, JX and JW wrote the paper.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Additional data can be accessed via the Dryad data repository at <http://datadryad.org/> with the doi:10.5061/dryad.s0s76.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- Bozgeyik Z, Coskun S, Dagli AF, *et al.* Diffusion-weighted MR imaging of thyroid nodules. *Neuroradiology* 2009;51:193–8.
- Erdem G, Erdem T, Muammer H, *et al.* Diffusion-weighted images differentiate benign from malignant thyroid nodules. *J Magn Reson Imaging* 2010;31:94–100.
- Koh DM, Collins DJ. Diffusion-weighted MRI in the body: applications and challenges in oncology. *AJR Am J Roentgenol* 2007;188:1622–35.
- Henzler T, Schmid-Bindert G, Schoenberg SO, *et al.* Diffusion and perfusion MRI of the lung and mediastinum. *Eur J Radiol* 2010;76:329–36.
- Elshafey R, Elattar A, Mlees M, *et al.* Role of quantitative diffusion-weighted MRI and 1H MR spectroscopy in distinguishing between benign and malignant thyroid nodules. *Egypt J Radiol Nucl Med* 2014;45:89–96.
- Wu Y, Yue X, Shen W, *et al.* Diagnostic value of diffusion-weighted MR imaging in thyroid disease: application in differentiating benign from malignant disease. *BMC Med Imaging* 2013;13:23.
- Shi HF, Feng Q, Qiang JW, *et al.* Utility of diffusion-weighted imaging in differentiating malignant from benign thyroid nodules with magnetic resonance imaging and pathologic correlation. *J Comput Assist Tomogr* 2013;37:505–10.
- Ilica AT, Artaş H, Ayan A, *et al.* Initial experience of 3 Tesla apparent diffusion coefficient values in differentiating benign and malignant thyroid nodules. *J Magn Reson Imaging* 2013;37:1077–82.
- Yue XH, Tao XF, Gao X. Application of diffusion-weighted MR imaging in the diagnosis of thyroid disease. *Chin J Radiol (China)* 2012;46:500–4.
- Nakahira M, Saito N, Murata S, *et al.* Quantitative diffusion-weighted magnetic resonance imaging as a powerful adjunct to fine needle aspiration cytology for assessment of thyroid nodules. *Am J Otolaryngol* 2012;33:408–16.
- Mutlu H, Sivrioglu AK, Sonmez G, *et al.* Role of apparent diffusion coefficient values and diffusion-weighted magnetic resonance imaging in differentiation between benign and malignant thyroid nodules. *Clin Imaging* 2012;36:1–7.
- El-Hariri MA, Gouhar Ghada K, Said Nagwa S, *et al.* Role of diffusion-weighted imaging with ADC mapping and in vivo 1H-MR spectroscopy in thyroid nodules. *Egypt J Radiol Nucl Med* 2012;43:183–92.
- Aydin H, Kizilgöz V, Tatar I, *et al.* The role of proton MR spectroscopy and apparent diffusion coefficient values in the diagnosis of malignant thyroid nodules: preliminary results. *Clin Imaging* 2012;36:323–33.
- Yan B, Liu HJ, Wang CB, *et al.* ADC values in differentiation of benign and malignant thyroid nodules. *Chin J Med Imaging Technol* 2011;27:510–14.
- Schueller-Weidekamm C, Kaserer K, Schueller G, *et al.* Can quantitative diffusion-weighted MR imaging differentiate benign and malignant cold thyroid nodules? Initial results in 25 patients. *AJNR Am J Neuroradiol* 2009;30:417–22.
- Ren S, Liu CH, Bai RJ. [Value of diffusion weighted imaging in diagnosis of nodular lesions of thyroid: a preliminary study]. *Zhonghua Yi Xue Za Zhi* 2010;90:3351–4.
- Li RK, QJ, Liu W, *et al.* Application of MR diffusion weighted imaging in the differentiation of malignant from benign thyroid focal lesions. *Radiol Practice (China)* 2009;24:719–22.
- Razek AA, Sadek AG, Kombar OR, *et al.* Role of apparent diffusion coefficient values in differentiation between malignant and benign solitary thyroid nodules. *AJNR Am J Neuroradiol* 2008;29:563–8.
- Whiting P, Rutjes AW, Reitsma JB, *et al.* 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, Weswood ME, Rutjes AW, *et al.* Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol* 2006;6:9.
- Whiting PF, Rutjes AW, Westwood ME, *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
- Higgins JP, Thompson SG, Deeks JJ, *et al.* Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557–60.
- Leeflang MM, Deeks JJ, Gatsonis C, *et al.* Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008;149:889–97.
- Zamora J, Abaira V, Muriel A, *et al.* Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* 2006;6:31.
- 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;12:1293–316.
- Reitsma JB, Glas AS, Rutjes AW, *et al.* Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982–90.
- Menke J. Bivariate random-effects meta-analysis of sensitivity and specificity with SAS PROC GLIMMIX. *Methods Inf Med* 2010;49:54–62, 62–4.
- StataCorp. *Stata statistical software: Release 10.* , 2009.
- 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;58:882–93.

30. Moher D, Liberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
31. Schueller-Weidekamm C, Schueller G, Kaserer K, *et al.* Diagnostic value of sonography, ultrasound-guided fine-needle aspiration cytology, and diffusion-weighted MRI in the characterization of cold thyroid nodules. *Eur J Radiol* 2010;73:538–44.
32. Jin Y, Qiang JW, Feng Q, *et al.* MR diffusion weighted imaging of thyroid nodules: correlation with pathology. *Chin J Med Imaging Technol* 2012;28:256–60.
33. Brown AM, Nagala S, McLean MA, *et al.* Multi-institutional validation of a novel textural analysis tool for reoperative stratification of suspected thyroid tumors on diffusion-weighted MRI. *Magn Reson Med* 2015. doi:10.1002/mrm.25743. [Epub ahead of print]
34. Vermoolen MA, Kwee TC, Nievelstein RA. Apparent diffusion coefficient measurements in the differentiation between benign and malignant lesions: a systematic review. *Insights Imaging* 2012;3:395–409.
35. Wu LM, Chen XX, Li YL, *et al.* On the utility of quantitative diffusion-weighted MR imaging as a tool in differentiation between malignant and benign thyroid nodules. *Acad Radiol* 2014;21:355–63.
36. Glas AS, Lijmer JG, Prins MH, *et al.* The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003;56:1129–35.
37. Pan Z, Trikalinos TA, Kavvoura FK, *et al.* Local literature bias in genetic epidemiology: an empirical evaluation of the Chinese literature. *PLoS Med* 2005;2:e334.
38. Vicker A, Goyal N, Harland R. Do certain countries produce only positive results? A systematic review of controlled trials. *Control Clin Trials* 1998;19:159–66.
39. Rutjes AW, Reitsma JB, Di Nisio M, *et al.* Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006;174:469–76.
40. Brazzelli M, Sandercock PA, Chappell FM, *et al.* Magnetic resonance imaging versus computed tomography for detection of acute vascular lesions in patients presenting with stroke symptoms. *Cochrane Database Syst Rev* 2009;(4):CD007424.
41. McAuley L, Pham B, Tugwell P, *et al.* Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet* 2000;356:1228–31.