REVIEW ARTICLE

The utility of MRI for pre-operative T and N staging of gastric carcinoma: a systematic review and meta-analysis

Z HUANG, MD, D H XIE, MD, PhD, L GUO, MD, C H HU, MD, PhD, X FANG, MD, Q MENG, MD, X X PING, MD and Z W LU, MD

Department of Radiology, The First Affiliated Hospital of Soochow University, Suzhou, China

Address correspondence to: Mr Dao Hai Xie
E-mail: *m18896586806@163.com*

Objective: To perform a meta-analysis and literature review regarding the diagnostic accuracy of MRI for pre-operative tumour depth invasion (T) and regional lymph node invasion (N) staging of gastric carcinoma (GC).

Methods: Articles were identified through systematic search of Medline, PubMed, Cochrane Library, Web of Science, Springerlink and several Chinese databases. The study quality was assessed by the quality assessment for studies of diagnostic accuracy. 2 reviewers independently extracted and assessed the data from 11 eligible studies. A meta-analysis was then carried out. Subgroup and sensitivity analyses were also performed.

Results: 11 studies (439 patients) were finally included in the current review. Among these studies, the significant evidence of heterogeneity was only discovered for specificity in T4 stage ($I^2 = 59.8\%$). Pooled sensitivity and specificity of MRI to diagnose T stage tumour (T3–4 *vs* T1–2) were 0.93 [95% confidence interval (CI), 0.89–0.96] and 0.91 (95% CI, 0.87–0.95), respectively. Pooled estimates of sensitivity and specificity of MRI to diagnose N stage tumour (N0 *vs* N+) were 0.86 (95% CI, 0.80–0.92) and 0.67 (95% CI, 0.54–0.79), respectively. Subgroup analyses showed that diffusion-weighted imaging was more helpful for T staging.

Conclusion: The present systematic review suggests that MRI has a good diagnostic accuracy for pre-operative T staging of GC and should be widely used in clinical work. However, the ability for N staging is relatively poor on MRI.

Advances in knowledge: In the pre-operative staging of GC, MRI was a useful tool and may enhance accuracy for the T staging of advanced GC.

Gastric carcinoma (GC) is the fourth most common cancer and the second leading cause of cancer-related death with a 5-year survival rate of <20% around the world.[1] The disease is more common in Asian countries, especially China, Japan and Republic of Korea.[2,3] Accurate assessment of local tumour depth invasion (T) and regional lymph node invasion (N) plays an essential role in predicting prognosis and determining the most appropriate treatment planning.[4,5]

The pre-operative staging of GC has been based on a multimodality approach, such as endoscopic ultrasonography (EUS), CT, MRI and positron emission tomography (PET).[6,7] EUS and CT have been widely used for GC staging in previous years.[8] Of course, different imaging modalities have themselves relative merits. CT with ionized radiation requires the injection of iodine contrast medium.[9] EUS is an invasive technique requiring sedation[1] and is highly operator dependent.[10] PET highly depends upon the standardized uptake value and the pathological subtype of the cancer.[11]

MRI is a powerful imaging method with high soft-tissue contrast, with technical versatility for sequence selection and modification, and without ionizing radiation. However, it was unsuitable for the staging of GC owing to its long acquisition time and susceptibility to motion artefacts in previous years. With technology improved and shorter imaging time, these limitations have recently been partially overcome.[12]

Recently, there has been much research using MRI to assess pre-operative staging of GC. Nevertheless, the number of patients in each study has been insufficient, and the results varied among the articles. Also, the limited imaging field of view of MRI in a single session makes it difficult to stage the distant metastasis (M).[13] Therefore, the objective of this study was to perform a systematic review and meta-analysis regarding the diagnostic accuracy of MRI for pre-operative T and N staging of GC.

METHODS AND MATERIALS
Literature search
A comprehensive computer literature search of studies on humans was performed. The Medline, PubMed, Cochrane

Library, Web of Science, Springerlink and several Chinese databases, including the Chinese Biomedical Literature Database, China National Knowledge Infrastructure and China Science and Technology Journal Database were searched with the following keywords: ("MRI" OR "magnetic resonance imaging") AND ("gastric cancer" OR "stomach cancer" OR "gastric adeno-carcinoma" OR "stomach adenocarcinoma" OR "gastric carcinoma" OR "stomach carcinoma") AND ("preoperative staging" OR "pre-operative staging" OR "preoperative imaging" OR "preoperative TNM staging" OR "diagnostic imaging"). We included the articles published before March 2014. Articles were limited in advance to the English and Chinese languages.

Inclusion and exclusion criteria

All electronic search titles, selected abstracts and full text of the obtained articles were read and assessed for inclusion independently by two reviewers (ZH and DHX) who were blinded to the journal, author, institution and date of publication. Disagreements were determined by consensus.

The inclusion criteria were: (a) studied pre-operative T and N staging performance of MRI in newly diagnosed patients with histopathology-confirmed GC; (b) included patients who underwent surgery, and pre-operative staging was compared with post-operative pathological staging; (c) included at least 30 patients who were suspected of GC; (d) only included the

Figure 1. Flowchart for identification of eligible studies. FN, false negative; FP, false positive; QUADS, quality assessment of diagnostic accuracy; TN, true negative; TP, true positive.
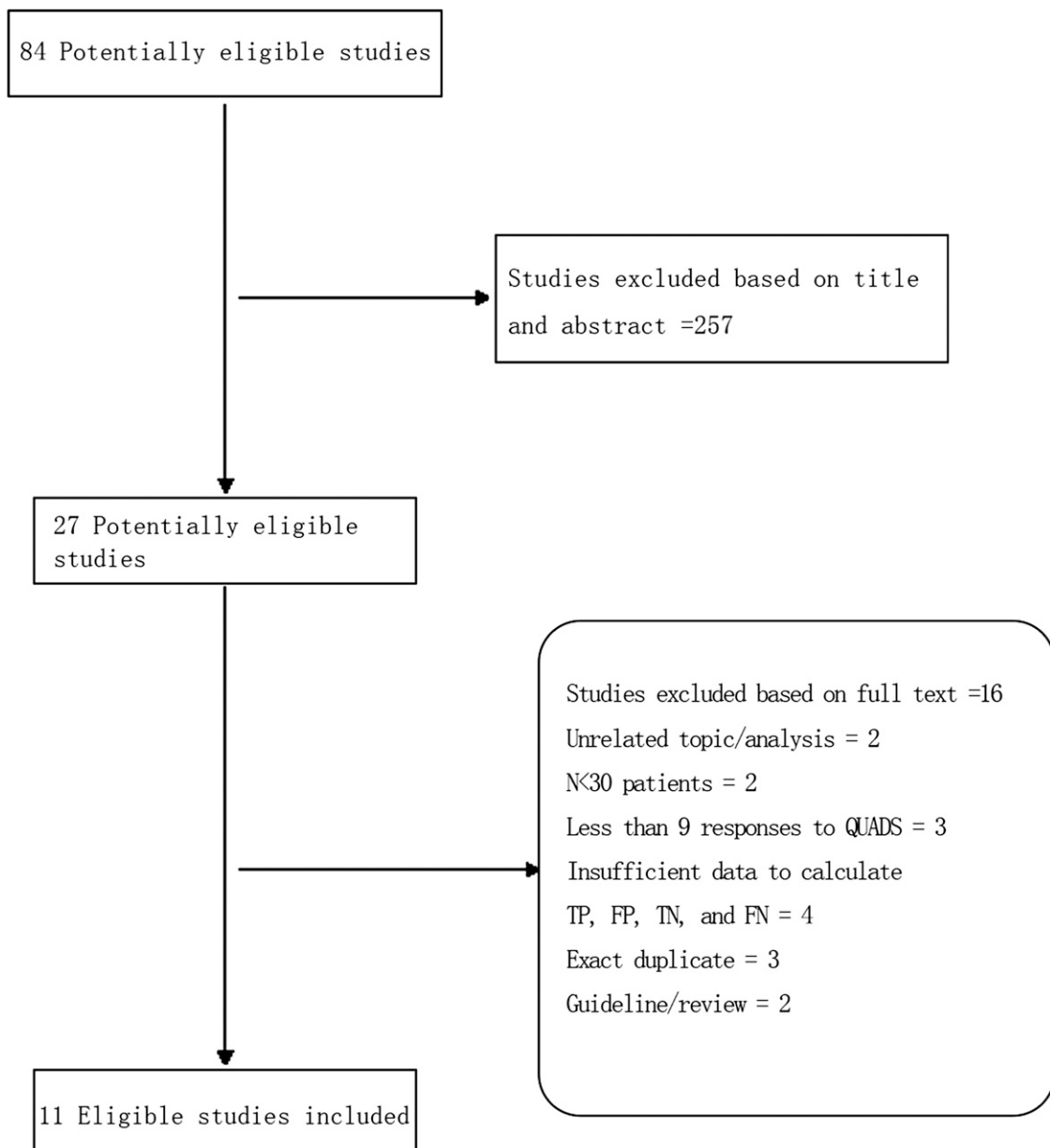
Table 1. Main characteristics of the included studies

Study, year	Country	Patients (n)	Mean age (years) (range)	Gender (M/F)	Blind	Sequence	b value (s mm^{-2})	Time of contrast (s)	Field strength (T)	Stage	T stage	T1/T2 (%)	N0 (%)	Study type	Delay time	Reference standards
Liu et al[9], 2014	China	51	M 64 (45–77); F 57 (28–82)	33/18	Y	T2, CE, DWI	0, 1000	30, 60, 90, 180	3.0 T	T	T1–T4	39	–	P	1–6 days	AJCC7
Joo et al[21], 2015	Republic of Korea	49	61.5 (38–81)	33/16	Y	T2, DWI, T1, CE	0, 100, 500, 10,000	18, 50, 90, 180	3.0 T	N	–	–	36	P	2–4 week	AJCC7
Lei et al[22], 2013	China	38	52 (31–82)	26/12	?	T1, T2, DWI, CE	600	30, 60, 90, 180	1.5 T	T, N	T1–T4	47	39	?	1 week	AJCC7
Tian et al[23], 2013	China	35	57.4 (35–74)	23/12	Y	T1, T2, CE	–	20, 50, 180	3.0 T	T	T2–T4	23	–	P	2–10 days	AJCC6
Huo et al[24], 2012	China	30	60 (50–69)	19/11	Y	T1, T2, DWI, CE	800, 1000	(25–30), (65–70), (180–240)	1.5 T	T, N	T1–T4	37	27	R	1 week	UICC6
Wang et al[25], 2011	China	42	54 (42–66)	18/24	?	T1, T2, DWI, CE	0, 600	20, 50, 180	1.5 T	T	T1–T4	21	–	R	?	UICC6
Anzidei et al[26], 2009	Italy	40	53.6	26/14	Y	T1, T2, CE	–	30, 60, 90	1.5 T	T	T0–T4	53	–	P	1 week	UICC/AJCC6
Wu et al[27], 2008	China	39	53 (31–75)	23/16	Y	T1, T2, CE	–	30, 60, 90	1.5 T	T, N	T1–T4	46	18	R	1 week	UICC/AJCC6
Li et al[28], 2007	China	35	56.2 (31–68)	23/12	Y	T2, T1, CE	–	30, 90, 180	1.5 T	T	T1–T4	49	–	?	1–7 days	AJCC4
Tang et al[29], 2006	China	35	57.8 (43–78)	25/10	Y	T1, T2, CE	–	27, (53–107), 180	0.5 T	T	T2–T4	23	–	P	1–6 days	UICC4
Kang et al[30], 2000	Republic of Korea	46	54 (26–69)	34/12	Y	T1, T2, CE	–	30, 60, 90	1.5 T	T, N	TX–T4	37	26	P	3–12 days	UICC4

AJCC, American Joint Committee on Cancer; CE, contrast enhanced; DWI, diffusion-weighted imaging; F, female; M, male; P, prospective; R, retrospective; UICC, Union for International Cancer Control; Y, yes.
? indicates that the corresponding information was not reported in the study.

Table 2. Evaluation of quality of included studies using the quality assessment of diagnostic accuracy studies tool

Study	Representative spectrum of patients	Selection criteria	Reference standard reliable	Time interval between MRI and pathology	Whole or random sample received verification	Same reference standard	Reference standard independent of the index test	Description of execution of MRI	Description of execution of pathology	Interpretation of MRI blinded from reference	Interpretation of reference blinded from MRI	Same clinical data available	Uninterruptable test results reported	Withdrawals explained
Liu et al[9]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y
Joo et al[21]	U	N	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y
Lei et al[22]	Y	N	Y	Y	Y	Y	Y	Y	N	U	Y	Y	Y	N
Tian et al[23]	Y	N	Y	Y	Y	Y	Y	Y	N	Y	U	Y	Y	Y
Huo et al[24]	Y	N	Y	Y	Y	Y	Y	Y	Y	U	U	Y	Y	Y
Wang et al[25]	Y	N	Y	U	Y	Y	Y	Y	Y	U	U	Y	Y	Y
Anzidei et al[26]	U	Y	Y	Y	Y	Y	Y	Y	N	Y	U	Y	Y	Y
Wu et al[27]	Y	N	Y	Y	Y	Y	Y	Y	N	Y	U	Y	Y	Y
Li et al[28]	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y
Tang et al[29]	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y
Kang et al[30]	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y

N, no; U, unclear; Y, yes.

articles published in English or Chinese; (e) reported sufficient data that could be labelled as true positive (TP), false positive (FP), true negative (TN) and false negative (FN) for T or/and N staging.

The exclusion criteria were as follows: (a) articles that investigated animals or/and *ex vivo* samples; (b) articles that did not provide enough information to determine pre-operative T and N staging performance; (c) studies did not include sufficient data to label as TP, FP, TN and FN for T or/and N staging; (d) studies with any pre-operative treatment or systematic chemotherapy; (e) review articles, meta-analyses, abstracts, conference proceedings, reports and letters.

Data extraction and quality assessment

Data were independently extracted by two reviewers (ZH and XF) from the obtained articles, including the first author's name, country, date of publication, study type, number of included patients, age, gender, MRI sequence and field strength, *b*-value, timing of imaging post-contrast injection, stage distribution for each of the study populations, delay times and the reference standards. Any controversy was resolved by consensus. The delay time was defined as the time interval between MRI and surgery.

The quality assessment of diagnostic accuracy studies (QUADAS) tool, which is an evidence-based quality assessment tool used in systematic reviews of diagnostic accuracy studies, was adopted to assess the methodological quality of the included articles.[14] Minimum criteria for fulfilling each QUADAS item were discussed by reviewers, and disagreements were resolved by consensus. It included 14 individual domains: representative spectrum of patients, selection criteria, reference standard reliable, time interval between MRI and pathology, whole or random sample received verification, same reference standard, reference standard independent of the index test, description of execution of MRI and pathology, interpretation of MRI blinded from reference, interpretation of reference blinded from MRI, same clinical data available, uninterruptable test results reported and withdrawals explained. Each item was assessed as "yes", "no" or "unclear".

Statistical analysis

The point estimates and their 95% confidence intervals (CIs) of the sensitivity, specificity, positive likelihood ratio (PLR) and negative likelihood ratio (NLR), and diagnostic odds ratio (DOR) were calculated from the data provided by the included studies in this review. If the PLR is >10, it suggests that the likelihood of correct staging of disease significantly increased; and if the NLR is <0.2, it suggests that the likelihood of correct staging of disease significantly decreased.[15] We also constructed the summary receiver-operating characteristic (SROC) curves, and reckoned the corresponding area under the curve representing the combination of the average sensitivity and specificity.

We used the Cochran Q test and I^2 statistic to assess the presence of heterogeneity in this review.[16] I^2 index represents the percent of total variation across all studies attributable to heterogeneity beyond chance; a higher value indicates more heterogeneity.[17] When the I^2 index was >50%, the random effects model was used to pool PLR, NLR and DOR values. When the I^2 index

Table 3. Summary estimates of diagnostic accuracy of MRI to diagnose T and N stages

Stage	Studies (n)	Patients (n)	Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)	Diagnostic odds ratio (95% CI)	Area under the curve	Heterogeneity (I^2, %)	Spearman correlation coefficient, p-value
T1	7	50	0.66 (0.51–0.79)	0.97 (0.94–0.99)	20.62 (9.49–44.80)	0.36 (0.25–0.52)	51.62 (20.75–128.44)	0.6737	0.0–29.0	−0.577, 0.175
T2	10	136	0.85 (0.78–0.90)	0.90 (0.85–0.93)	6.11 (4.18–8.91)	0.21 (0.14–0.30)	36.03 (18.99–68.37)	0.9219	0.0–21.6	0.639, 0.053
T3	8	119	0.86 (0.78–0.91)	0.89 (0.83–0.93)	6.69 (4.43–10.11)	0.18 (0.12–0.27)	39.99 (20.38–78.47)	0.9254	0.0–32.9	0.180, 0.670
T4	8	73	0.88 (0.78–0.94)	0.97 (0.94–0.99)	16.57 (7.29–37.65)	0.20 (0.12–0.33)	106.71 (38.53–295.54)	0.9511	0.0–59.8	−0.054, 0.899
T3–4	10	216	0.93 (0.89–0.96)	0.91 (0.87–0.95)	7.66 (4.87–12.03)	0.10 (0.06–0.16)	92.12 (42.63–199.08)	0.9622	0.0–37.7	0.052, 0.097
N	5	201	0.86 (0.80–0.92)	0.67 (0.54–0.79)	2.59 (1.80–3.73)	0.21 (0.13–0.33)	12.75 (6.31–25.77)	0.8848	0.0–21.6	0.400, 0.505

CI, confidence interval.

was <50%, we used the fixed effects model to pool these values.[18] Publication bias was visually assessed by funnel plots. An asymmetric funnel shape represents a significant bias. A regression of the logarithm of DOR (lnDOR) with 1/effective sample size$^{1/2}$ was used to calculate the degree of asymmetry. For the slope coefficient, $p < 0.10$ indicates significant asymmetry of the funnel plots.[19]

The SROC curve was also used to visually assess the between-study variation caused by the threshold effect. And the variation was calculated as the squared coefficient of correlation between logit sensitivity and specificity estimated by the bivariate model. To investigate the influence of individual quality items, we conducted sensitivity analyses by excluding retrospective investigations, studies with unclear or high risk of bias in the patient selection and outliers.[20] A subgroup analysis was performed to compare the diagnostic performance with diffusion-weighted imaging (DWI) to those without DWI for T staging of GC.

Two statistical software packages (Meta-Disc v. 1.4 and Stata®/ SE v. 12.0; Stata Corporation, College Station, TX) were used to perform the statistical analysis.

RESULTS

Literature search and selection of studies

A flow chart of the inclusion of studies in the meta-analysis is shown in Figure 1. Finally, 11 eligible studies published between 2000 and 2015[9,21–30] fulfilled all the inclusion criteria and were included in our review.

Study characteristics

The principal characteristics of the 11 studies included in this meta-analysis are showed in Table 1. There were a total of 439 patients enrolled, with a predominance of males and a mean of 40 patients per study (range, 30–51 patients). In these articles, the majority (10 articles, 400 patients) of the studies were from Asia. In 11 studies, the age ranged from 31 to 82 years. Six studies enrolled patients prospectively, three studies were retrospective and the other two did not define the status. 9 of the 11 studies used blind method, but the remaining 2 did not report this information. Blind method was defined as the investigators assessing the MRI images without knowledge of the pathology results.

Quality assessment

The quality of the 11 qualified studies, as assessed according to the QUADAS criteria, is reported in Table 2. It involved 14 items. Six of them could be scored in all selected articles. Common disadvantages were concentrated in Items 2 and 11. Selection criteria (Item 2) was not present in 81.8% (9) of the 11 articles. The blind measurements of pathology without knowledge of MRI (Item 11) were not present in 90.9% (10) of the 11 articles. Representative spectrum (Item 1) was not present in 18.2% (2) of the 11 articles. The execution of the reference standard in detail (Item 9) was not present in 36.4% (4) of the 11 articles. The measurements of MRI staging without knowledge of the results of the reference standard (Item 10) were not present in 27.3% (3) of the 11 articles.

Figure 2. Forest plots of pooled sensitivity of MRI to diagnose T stage. CI, confidence interval; df, degrees of freedom.
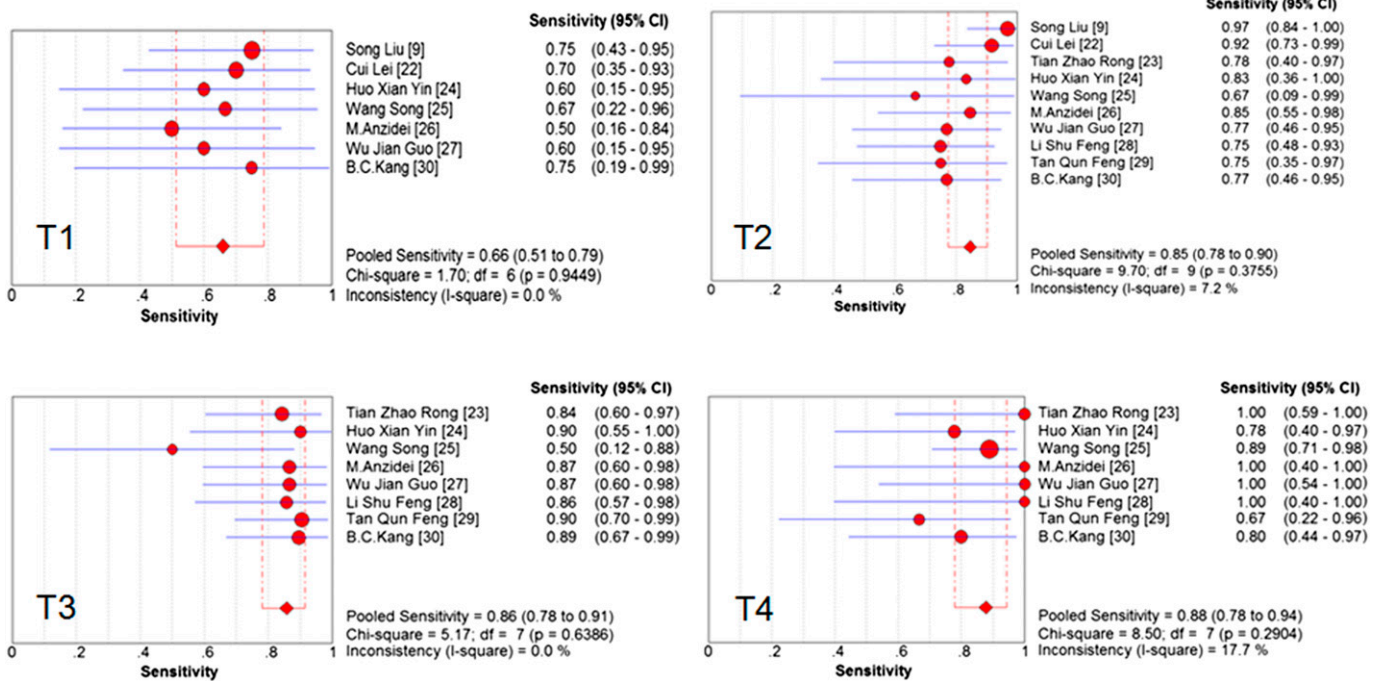

Diagnostic accuracy of primary tumour depth (T stage)

This meta-analysis article included various articles using the seventh, sixth and fourth staging system of the Union for International Cancer Control (UICC) and/or the American Joint Committee on Cancer (AJCC) staging classifications. There are too many changes of T-staging according to the change of edition. For example, T2 in the sixth edition is T2 and T3 in the seventh edition; and T3 and T4 in the sixth edition is T4 in the seventh edition, respectively. The T2 stage also has slight distinction among the sixth and fourth editions of the UICC/AJCC system; for the purposes of this meta-analysis, there are no differences among the sixth and fourth editions.[6] To make the various editions consistent in our review, we used the sixth and fourth editions as the reference standards. T2 and T3 in the seventh edition were regrouped into T2, and T4 in the seventh

Figure 3. Forest plots of pooled specificity of MRI to diagnose T stage. CI, confidence interval; df, degrees of freedom.
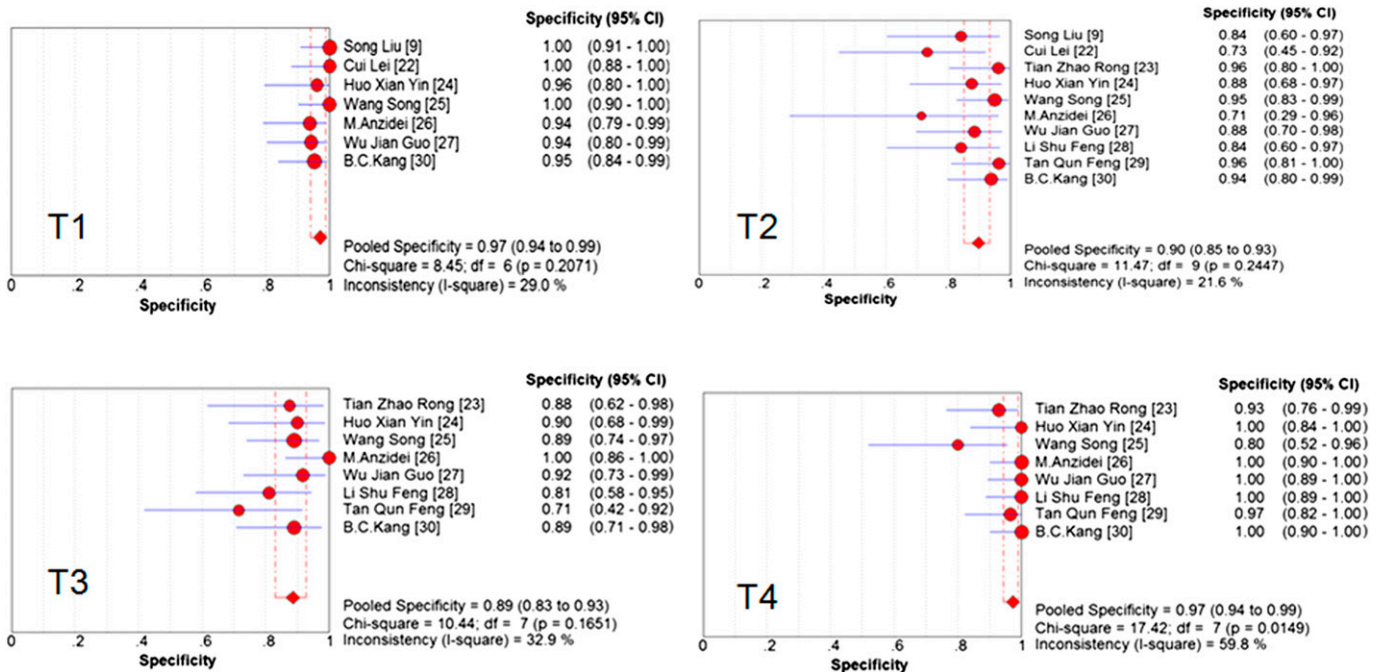
edition was also regrouped into T3–4. The number of the patients confirmed pathologically to be Stages T1, T2, T3 and T4 is shown in Table 3, respectively.

The pooled accuracy of MRI to diagnose T stage tumour was 0.81 (95% CI, 0.77–0.84). To assess the diagnostic accuracy of MRI to separate early-to-intermediate (T1–2) and even advanced (T3–4) GC, we performed a meta-analysis of the included studies on T staging (T3–4 *vs* T1–2). For this purpose, there were 10 studies (213 patients) available, and the sample size of $n = 213$ referred to the patients with T3 and T4 staging in this review. Pooled sensitivity and specificity were 0.93 (95% CI, 0.89–0.96) and 0.91 (95% CI, 0.89–0.95), respectively. The sensitivity and specificity of MRI to stage T1, T2, T3, T4 and T3–4 stage tumours are shown as forest plots in Figures 2–4. The PLR and DOR for various T stages are displayed in Table 3. The SROC curves for T1, T2, T3, T4 and T3–4 stages are also shown in Figures 4 and 5.
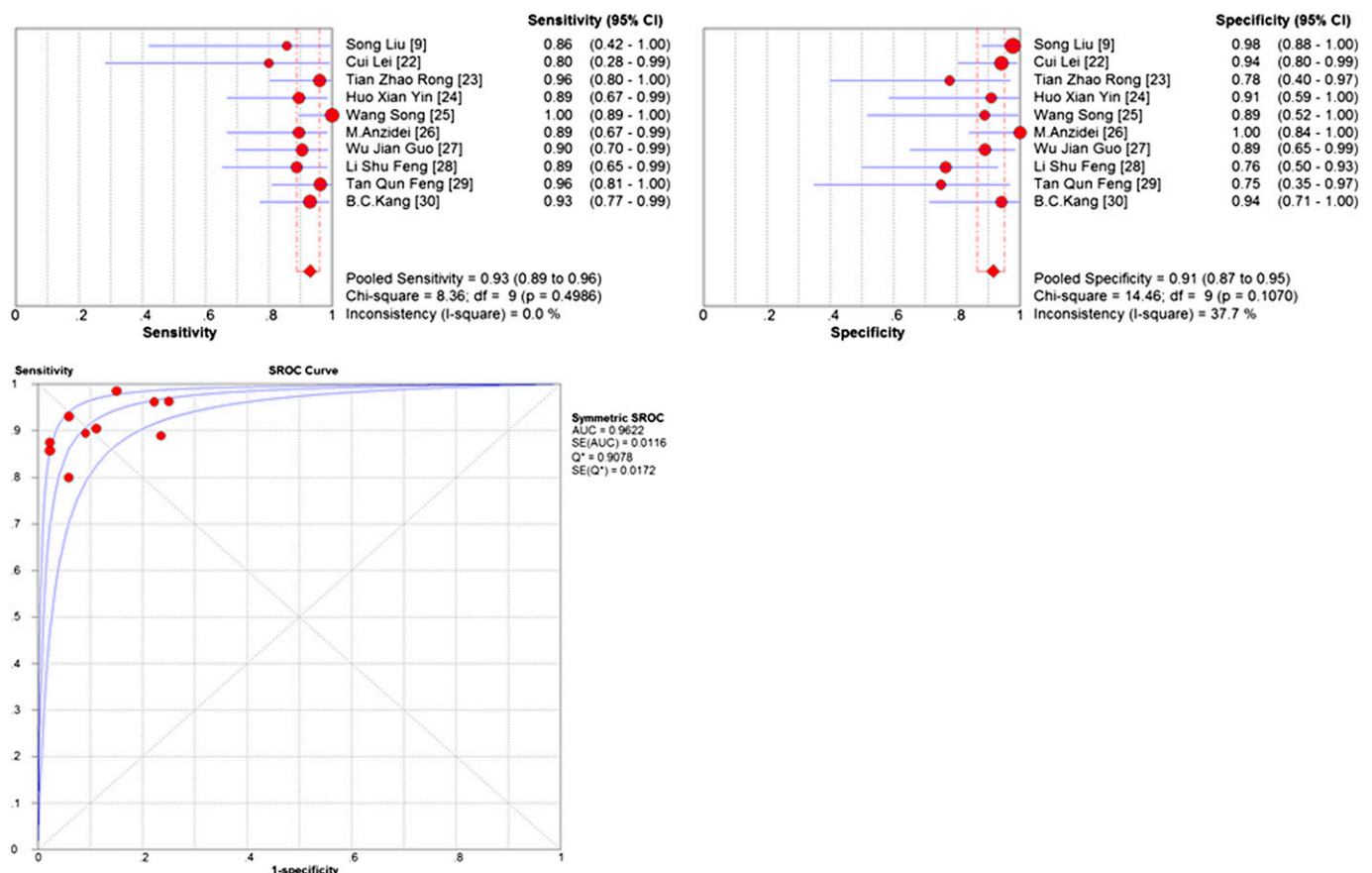
The Cochran Q test (Figure 3) confirmed that the significant evidence of heterogeneity was only found for specificity in T4 stage tumour ($I^2 = 59.8\%$) in the included studies. Therefore we pooled the estimates of MRI for T1, T2, T3 and T3–4 staging in the fixed effects model and for T4 staging in the random effects model. The ratio of heterogeneity probably caused by the threshold effect was very low (Table 3). There was only one outlying study with a low specificity of MRI for

T4 staging.[21] A sensitivity analysis was performed to investigate the influence of excluding the study on MRI specificity for T4 staging. We found that this analysis produced a higher estimate of specificity, and statistical heterogeneity of specificity in T4 stage tumour was diminished. For T4 stage, we also performed a sensitivity analysis to investigate the influence of excluding the retrospective studies. We found that there was no statistically significant difference between the different study types.

Diagnostic accuracy of lymph node status (N stage)

The classifications of N stage are also different among the various editions of UICC/AJCC. In the UICC/AJCC system, N stage is defined according to the distance of the perigastric regional lymph nodes from the edge of the primary tumour or the total number of lymph node metastases present.[31,32]

Our meta-analysis only compared the performance of MRI to identify N0 *vs* N+ disease. Thus, 5 articles (201 patients) were included. The pooled accuracy of MRI to diagnose N stage tumour was 0.78 (95% CI, 0.72–0.83). Summary estimates of sensitivity, specificity, DOR, PLR and NLR were 0.86 (95% CI, 0.80–0.92), 0.67 (95% CI, 0.54–0.79), 2.59 (95% CI, 1.80–3.73), 0.21 (95% CI, 0.13–0.33) and 12.75 (95% CI, 6.31–25.77), respectively. The sensitivity and specificity of MRI for N stage are presented as Forrest plots (Figure 6). The AUC curve for N stage is also presented in Table 3.

Figure 4. Pooled sensitivity, specificity, and summary receiver-operating characteristic (SROC) of MRI to diagnose T3–4 stage. AUC, area under the curve; CI, confidence interval; SE, standard error.

The Cochran Q test (Figure 6) confirmed that there was little heterogeneity across the studies for N stage.

Subgroup analyses

We compared the pooled T (T3–4 *vs* T1–2) performance characteristics of MRI with DWI[9,22,24,25] to those without DWI[23,26–30] to determine whether DWI could have helped pre-operative staging performances (Table 4). However, we found that the summary results for T staging of GC (Table 4) showed no statistically significant difference between MRI with DWI and without DWI in sensitivity ($p = 0.279$) and specificity ($p = 0.283$). There were not enough studies available to perform subgroup analyses for N staging in our review.

Publication bias

In this meta-analysis, the bias calculations and funnel plots did not show statistically significant bias for T2 ($p = 0.36$), T3 ($p = 0.11$), T4 ($p = 0.12$) and T3–4 ($p = 0.52$) staging. However, there were insufficient included studies to allow assessment of reporting bias for T1 ($n = 7$) and N ($n = 5$) staging. Therefore, we concluded that publication bias may exist in T1 and N stages.

DISCUSSION

It is widely known that accurate pre-operative staging can improve therapeutic approaches and prognosis for patients with GC.[33,34] Recently, MRI as a helpful imaging tool for GC staging has been widely used worldwide. In order to identify 11 studies that investigated the diagnostic accuracy of MRI for pre-operative staging of GC, we searched nearly 284 titles and abstracts in this systematic review. To the best of our knowledge, this is the first meta-analysis that singularly quantitatively summarizes the diagnostic performance of MRI in the loco-regional staging of GC to include a large cohort of patients ($n = 439$).

A systematic review by Seevaratnam et al[6] evaluated 40 articles, of which 3 evaluated GC with MRI ($n = 109$) and showed that

Figure 5. Summary receiver-operating characteristic (SROC) of MRI to diagnose T stage. AUC, area under the curve; df, degrees of freedom; SE, standard error.
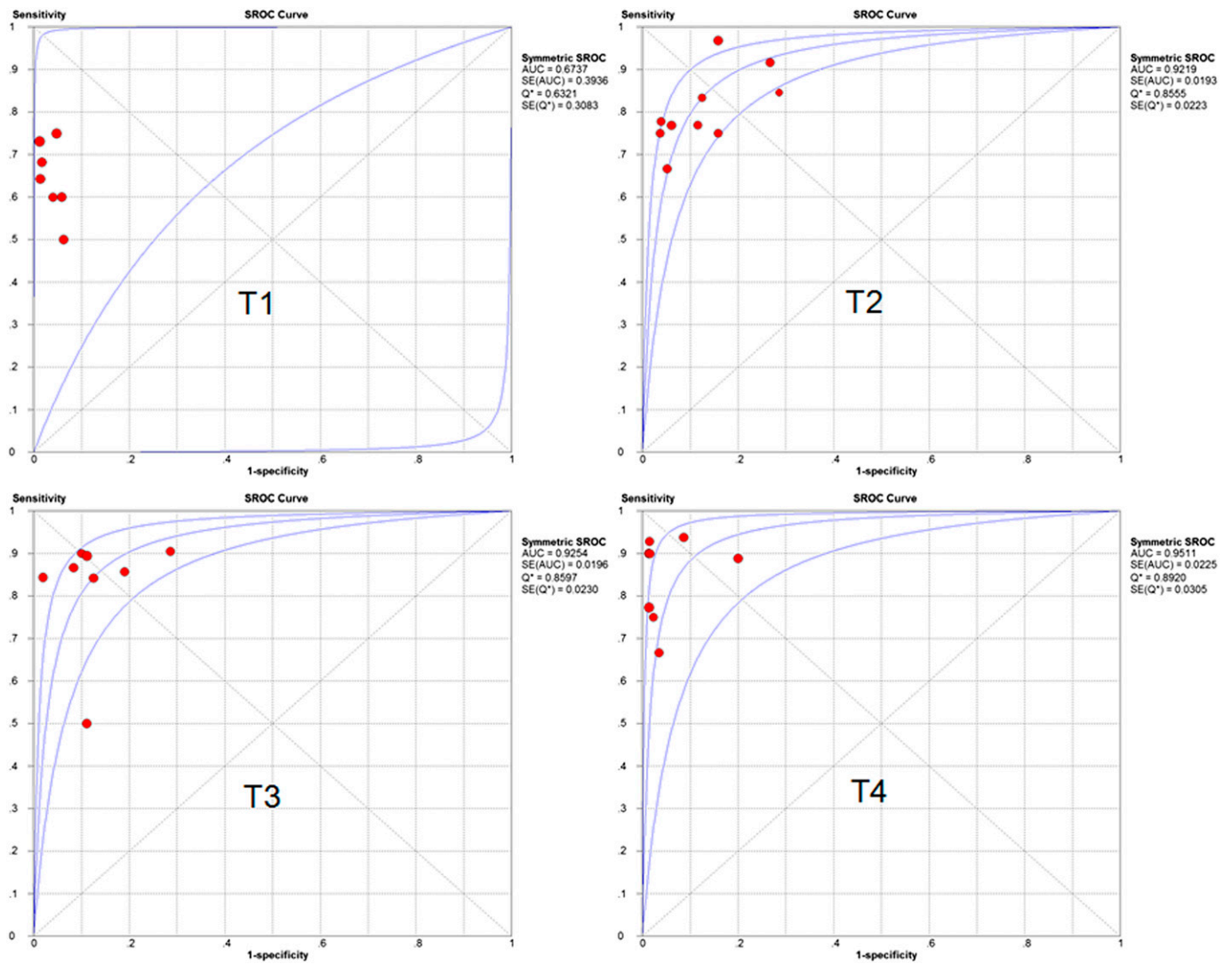
MRI had the best overall performance characteristics for tumour invasion (T) staging compared with other imaging modalities such as abdominal ultrasound, CT and PET. In this review, it was found that the overall accuracy for T stage was 68%, 72% and 83% and the overall accuracy for N stage was 68%, 66% and 53% for abdominal ultrasound, CT and MRI, respectively. However, what should be noted is that only three MRI studies published in 2000 were included in their review. What we have concluded is similar to the findings of Seevaratnam et al, although our meta-analysis found that MRI had a weakness for T1 staging. There were two systematic reviews assessing diagnostic accuracy of EUS for the staging of GC in previous years. Cardoso et al[35] found that EUS was a moderately accurate technique for both T and N staging. Mocellin et al[36] found that the sensitivity of the single T categories (T1, T2, T3 and T4) for EUS were 83%, 65%, 86% and 66%, respectively, and the specificity were 96%, 91%, 85% and 98%, respectively. Compared with the previous reviews, we found that MRI performed better for T staging except for T1 stage. An explanation for the sufficient diagnostic performance of MRI in T2–4 staging is its excellent soft-tissue contrast. The limitation of MRI for T1 staging probably may be related to the location of the lesions, such as tumour in the angle or cardia of the stomach, which is easily underestimated for the native thickness of the gastric wall, and histological differentiation of carcinoma with lower grades of enhancement that is hardly detected in the process of enhancement.[32] Thus, we need further investigations with large sample sizes to confirm our results.

In this review, our results showed that the pooled sensitivity of MRI for T staging ranged from 66% to 93%, with a relatively higher value (93%) observed in T3–4 lesions than in T1–2 lesions (Table 3). For the overall T stages, the pooled specificity ranged from 89% to 97%. And the overall accuracy of MRI for T stage was 81%. The pooled accuracy, sensitivity and specificity for detecting nodal invasion were 78%, 86% and 67%, respectively. In conclusion, we found that MRI performed better for T staging than for N staging, especially for specificity. Kwee and Kwee[1] also came to a conclusion that no one imaging tool consistently attains both high sensitivity and high specificity in N staging of GC. Lymph node size and enhancing lymph node status were used by all studies included in the review as the criteria to define malignancy.[29,30] The inability to identify metastatic lymph node status with normal size may explain the insufficient diagnostic accuracy of MRI. Although the diagnostic performance of MRI for N staging is not better than T staging, this imaging modality can also help clinicians to consider the risk of patients with lymph node metastatic disease to select suitable pre-operative treatments.

The PLR is an indicator to what extent the test could identify a disease, and the NLR is a measure of how well the same test performs in excluding the disease.[13] In our review, We found that MRI had the lowest NLR for T3–4 staging, suggesting that the negative examination results of MRI for GC staging could be used alone as a justification to rule out the T3–4 stage of the

Figure 6. Pooled sensitivity, specificity and summary receiver-operating characteristic (SROC) of MRI to diagnose N stage. AUC, area under the curve; CI, confidence interval; df, degrees of freedom; SE, standard error.
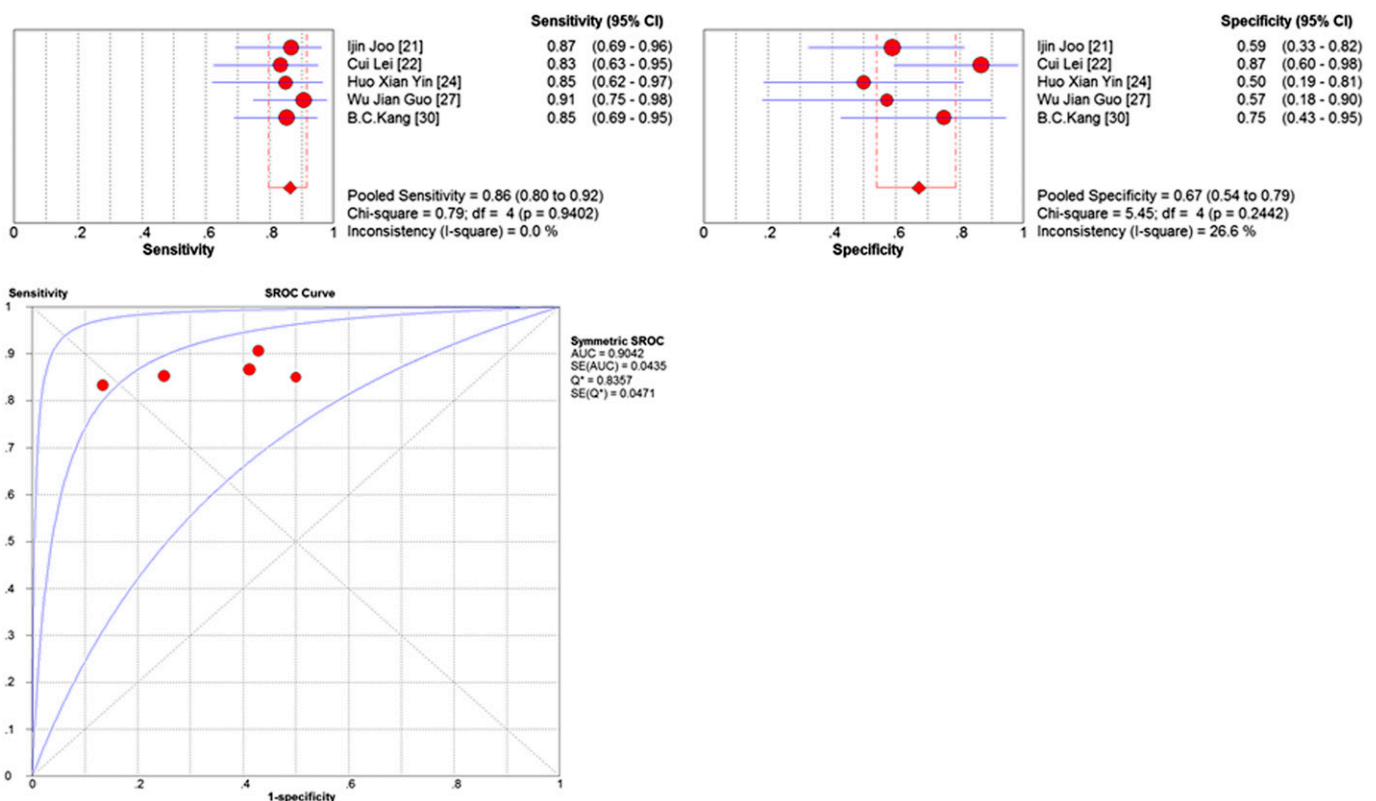
Table 4. Comparison of MRI performance characteristics to diagnose T3–4 stages with and without diffusion-weighted imaging (DWI)

MRI	Studies (*n*)	Patients (*n*)	Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)	Diagnostic odds ratio (95% CI)	Area under the curve	Heterogeneity (I^2, %)
With DWI	4	101	0.94 (0.85–0.98)	0.95 (0.88–1.00)	10.30 (4.30–24.69)	0.11 (0.05–0.26)	132.46 (33.74–520.01)	0.9722	0.0–53.1
Without DWI	6	145	0.93 (0.87–0.97)	0.88 (0.79–0.94)	6.83 (4.02–11.62)	0.10 (0.05–0.17)	80.10 (31.83–201.58)	0.9629	0.0–47.9

CI, confidence interval.

disease. We also found that MRI had the highest PLR for T1 staging of GC in this analysis, indicating that MRI is a sensitive measure to identify the T1 stage of this disease, although T1 staging had the lowest sensitivity.

DOR is defined as the odds of having a positive test in patients with a true anatomical stage of the disease when compared with patients who do not have the disease.[37] DOR has a value that ranges from zero to infinity, with higher value indicating better discriminatory test performance.[38] The results of this meta-analysis indicated that MRI had an effective diagnostic performance in T staging of GC, especially for T4 and T3–4 stage. An area under the curve of one for any diagnostic test indicates that the test is extremely accurate.[39] In our findings, except for T1 stage disease, SROC curves of MRI for the pre-operative staging of GC were close to one, showing that MRI is a reliable test for advanced GC staging.

In our review, statistical heterogeneity was only found in T4 stage specificity estimates and appeared to reduce when one study with low specificity was excluded in a sensitivity analysis, suggesting that the variability could be explained by the outlying point alone. The threshold effect was not found in our meta-analysis. Subgroup analyses were also performed to further investigate the different diagnostic performances between MRI with DWI and without DWI for T stage. We found that the results did not show any statistically significant difference between them. Although DWI may be beneficial for detecting lymph node metastases, testified by several studies,[40] there were not enough studies available to perform subgroup analyses for N staging in our review. Therefore, further research is needed to explore the DWI for N staging of GC.

Our study was based on thorough literature search and careful data extraction, and included assessments of the methodological quality of diagnostic test accuracy studies. However, some limitations should be mentioned here. First of all, publication bias could not be avoided in our review for the small size of our included studies. Second, the majority of the included patients in our meta-analysis were from Asia. Accordingly, the results may not be useful in other regions. Third, the inclusion bias should be considered in our study, because only the studies published in English or Chinese were selected. Fourth, six studies were prospectively designed, three studies retrospectively designed and two studies did not specify their design, which resulted in a selection bias in this review. Fifth, the proportion of patients with different T stages varied between studies, which may cause a bias when evaluating the overall diagnostic performance of MRI.

CONCLUSION

The present meta-analysis shows that MRI has a good diagnostic performance for the T staging of advanced GC, especially for T3–4 staging. Although T1 staging has the lowest sensitivity, MRI is also a sensitive measure to identify the T1 stage of this disease for the highest PLR value. But for N staging, as like other imaging modalities (EUS, CT, PET-CT) reported by previous reviews,[32–34] MRI has a poor diagnostic

performance in GC staging. DWI may be beneficial for detecting lymph node metastases, which was confirmed by several studies. Therefore, researchers should pay more attention to DWI for N staging in the future.

FUNDING

REFERENCES

1. Kwee RM, Kwee TC. Imaging in assessing lymph node status in gastric cancer. *Gastric Cancer* 2009; **12**: 6–22. doi: 10.1007/s10120-008-0492-5

2. Moore MA, Manan AA, Chow KY, Cornain SF, Devi CR, Triningsih FX, et al. Cancer epidemiology and control in peninsular and island South-East Asia—past, present and future. *Asian Pac J Cancer Prev* 2010; **11**: 81–98.

3. Kamangar F, Dores GM, Anderson WF. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J Clin Oncol* 2006; **24**: 2137–50. doi: 10.1200/jco.2005.05.2308

4. Cunningham D, Allum WH, Stenning SP, Thompson JN, Van de Velde CJ, Nicolson M, et al; MAGIC Trial Participants. Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. *N Engl J Med* 2006; **355**: 11–20. doi: 10.1056/NEJMoa055531

5. Yoshikawa T, Sasako M, Yamamoto S, Sano T, Imamura H, Fujitani K, et al. Phase II study of neoadjuvant chemotherapy and extended surgery for locally advanced gastric cancer. *Br J Surg* 2009; **96**: 1015–22. doi: 10.1002/bjs.6665

6. Seevaratnam R, Cardoso R, McGregor C, Lourenco L, Mahar A, Sutradhar R, et al. How useful is preoperative imaging for tumor, node, metastasis (TNM) staging of gastric cancer? A meta-analysis. *Gastric Cancer* 2012; **15**: 3–18. doi: 10.1007/s10120-011-0069-6

7. Yin XD, Huang WB, Lü CY, Zhang L, Wang LW, Xie GH. A preliminary study on correlations of triple-phase multi-slice CT scan with histological differentiation and intratumoral microvascular/lymphatic invasion in gastric cancer. *Chin Med J (Engl)* 2011; **124**: 347–51.

8. Hwang SW, Lee DH, Lee SH, Park YS, Hwang JH, Kim JW, et al. Preoperative staging of gastric cancer by endoscopic ultrasonography and multidetector-row computed tomography. *J Gastroenterol Hepatol* 2010; **25**: 512–18. doi: 10.1111/j.1440-1746.2009.06106.x

9. Liu S, He J, Guan W, Li Q, Yu H, Zhou Z, et al. Added value of diffusion-weighted MR imaging to T2-weighted and dynamic contrast-enhanced MR imaging in T staging of gastric cancer. *Clin Imaging* 2014; **38**: 122–8. doi: 10.1016/j.clinimag.2013.12.001

10. Pollack BJ, Chak A, Sivak MV Jr. Endoscopic ultrasonography. *Semin Oncol* 1996; **23**: 336–46.

11. Dassen AE, Lips DJ, Hoekstra CJ, Pruijt JF, Bosscha K. FDG-PET has no definite role in preoperative imaging in gastric cancer. *Eur J Surg Oncol* 2009; **35**: 449–55. doi: 10.1016/j.ejso.2008.11.010

12. Heye T, Kuntz C, Düx M, Encke J, Palmowski M, Autschbach F, et al. CT and endoscopic ultrasound in comparison to endoluminal MRI: preliminary results in staging gastric carcinoma. *Eur J Radiol* 2009; **70**: 336–41. doi: 10.1016/j.ejrad.2008.01.037

13. Koh DM, Collins DJ. Diffusion-weighted MRI in the body: applications and challenges in oncology. *AJR Am J Roentgenol* 2007; **188**: 1622–35. doi: 10.2214/AJR.06.1403

14. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; **3**: 25. doi: 10.1186/1471-2288-3-25

15. Liao CY, Chen JH, Liang JA, Yeh JJ, Kao CH. Meta-analysis study of lymph node staging by 18 F-FDG PET/CT scan in non-small cell lung cancer: comparison of TB and non-TB endemic regions. *Eur J Radiol* 2012; **81**: 3518–23. doi: 10.1016/j.ejrad.2012.02.007

16. Dinnes J, Deeks J, Kirby J, Roderick P. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. *Health Technol Assess* 2005; **9**: 1–113. doi: 10.3310/hta9120

17. Huedo-Medina TB, Sánchez-Meca J, Marin-Martinez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychol Methods* 2006; **11**: 193–206. doi: 10.1037/1082-989X.11.2.193

18. DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. *Contemp Clin Trials* 2007; **28**: 105–14. doi: 10.1016/j.cct.2006.04.004

19. 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. doi: 10.1016/j.jclinepi.2005.01.016

20. Al-Sukhni E, Milot L, Fruitman M, Beyene J, Victor JC, Schmocker S, et al. Diagnostic accuracy of MRI for assessment of T category, lymph node metastases, and circumferential resection margin involvement in patients with rectal cancer: a systematic review and meta-analysis. *Ann Surg Oncol* 2012; **19**: 2212–23. doi: 10.1245/s10434-011-2210-5

21. Joo I, Lee JM, Kim JH, Shin CI, Han JK, Choi BI, et al. Prospective comparison of 3T MRI with diffusion-weighted imaging and MDCT for the preoperative TNM staging of gastric cancer. *J Magn Reson Imaging* 2015; **41**: 814–21. doi: 10.1002/jmri.24586

22. Lei C, Huang L, Wang Y, Huang Y, Huang Y. Comparison of MRI and endoscope ultrasound detection in preoperative T/N staging of gastric cancer. *Mol Clin Oncol* 2013; **1**: 699–702. doi: 10.3892/mco.2013.103

23. Tian ZR, Guo YL, Zhu K, Niu JD. Application of 3.0 T MRI in preoperative T staging of advanced gastric cancer. *Chin J Med Imaging Technol* 2013; **29**: 433–6.

24. Huo XY, Yuan KS, Yang J, Lu LH, Zhai X, Ke XL, et al. Comparative study on NM staging of gastric cancer between magnetic resonance imaging (MRI) detection preoperative and pathological diagnosis postoperative. *J Chin Physician* 2012; **zl**: 95–7.

25. Wang S, Ren K, Sun WG, Wang Q, Han M. Comparative study of MRI and MDCT in the T staging of gastric carcinoma be surgery. *Radiol Practice* 2011; **26**: 426–9.

26. Anzidei M, Napoli A, Zaccagna F, Di Paolo P, Zini C, Cavallo Marincola MB, et al. Diagnostic performance of 64-MDCT and 1.5-T MRI with high-resolution sequences in the T staging of gastric cancer: a comparative analysis with histopathology. *Radiol Med* 2009; **114**: 1065–79. doi: 10.1007/s11547-009-0455-x

27. Wu JG, Fang GE, Luo TH, Zhang F. Value of dynamic subtraction technique of magnetic resonance imaging in preoperative TNM-staging assessment of gastric carcinoma. [In Chinese.] *Chin J Gastrointest Surg* 2008; **11**: 533–6.

28. Li SF, Zhao RF, Li HB, Li JD, Jin JL. Dynamic contrast enhanced MR study on preoperative TNM staging of advanced gastric carcinoma.

Chin J Med Imaging Technol 2007; **23**: 1187–90.

29. Tang QF, Shen JK, Feng YZ, Li JD, Jin JL. Evaluation of dynamic 0.5T MRI in the preoperative TNM-staging of advanced gastric carcinoma. *Chin J Med Imaging Technol* 2006; **22**: 88–91.

30. Kang BC, Kim JH, Kim KW, Lee DY, Baek SY, Lee SW, et al. Value of the dynamic and delayed MR sequence with Gd-DTPA in the T-staging of stomach cancer: correlation with the histopathology. *Abdom Imaging* 2000; **25**: 14–24. doi: 10.1007/s002619910003

31. Yan C, Zhu ZG, Yan M, Zhang H, Pan ZL, Chen J, et al. Value of multidetector-row computed tomography in the preoperative T and N staging of gastric carcinoma: a large-scale Chinese study. *J Surg Oncol* 2009; **100**: 205–14. doi: 10.1002/jso.21316

32. Ahn HS, Lee HJ, Yoo MW, Kim SG, Im JP, Kim SH, et al. Diagnostic accuracy of T and N stages with endoscopy, stomach protocol CT, and endoscopic ultrasonography in early

gastric cancer. *J Surg Oncol* 2009; **99**: 20–7. doi: 10.1002/jso.21170

33. Li B, Zheng P, Zhu Q, Lin J. Accurate preoperative staging of gastric cancer with combined endoscopic ultrasonography and PET-CT. *Tohoku J Exp Med* 2012; **228**: 9–16. doi: 10.1620/tjem.228.9

34. Chen CY, Hsu JS, Wu DC, Kang WY, Hsieh JS, Jaw TS, et al. Gastric cancer: preoperative local staging with 3D multi-detector row CT—correlation with surgical and histopathologic results. *Radiology* 2007; **242**: 472–82. doi: 10.1148/radiol.2422051557

35. Cardoso R, Coburn N, Seevaratnam R, Sutradhar R, Lourenco LG, Mahar A, et al. A systematic review and meta-analysis of the utility of EUS for preoperative staging for gastric cancer. *Gastric Cancer* 2012; **15**: S19–26. doi: 10.1007/s10120-011-0115-4

36. Mocellin S, Marchet A, Nitti D. EUS for the staging of gastric cancer: a meta-analysis. *Gastrointest Endosc* 2011; **73**: 1122–34. doi: 10.1016/j.gie.2011.01.030

37. Plana MN, Carreira C, Muriel A, Chiva M, Abraira V, Emparanza JI, et al. Magnetic resonance imaging in the preoperative assessment of patients with primary breast cancer: systematic review of diagnostic accuracy and meta-analysis. *Eur Radiol* 2012; **22**: 26–38. doi: 10.1007/s00330-011-2238-8

38. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003; **56**: 1129–35. doi: 10.1016/s0895-4356(03)00177-x

39. Trikudanathan G, Njei B, Attam R, Arain M, Shaukat A. Staging accuracy of ampullary tumors by endoscopic ultrasound: meta-analysis and systematic review. *Dig Endosc* 2014; **26**: 617–26. doi: 10.1111/den.12234

40. Akduman EI, Momtahen AJ, Balci NC, Mahajann N, Havlioglu N, Wolverson MK. Comparison between malignant and benign abdominal lymph nodes on diffusion-weighted imaging. *Acad Radiol* 2008; **15**: 641–6. doi: 10.1016/j.acra.2007.12.023