

How useful is preoperative imaging for tumor, node, metastasis (TNM) staging of gastric cancer? A meta-analysis

Rajini Seevaratnam · Roberta Cardoso · Caitlin McGregor · Laercio Lourenco · Alyson Mahar · Rinku Sutradhar · Calvin Law · Lawrence Paszat · Natalie Coburn

Received: 28 January 2011 / Accepted: 29 May 2011 / Published online: 12 August 2011
© The International Gastric Cancer Association and The Japanese Gastric Cancer Association 2011

Abstract

Background Surgery is the fundamental curative option for gastric cancer patients. Imaging scans are routinely prescribed in an attempt to stage the disease prior to surgery. Consequently, the correlation between radiology exams and pathology is crucial for appropriate treatment planning.

Methods Systematic searches were conducted using Medline, Embase, and the Cochrane Central Register of Controlled Trials from January 1, 1998 to December 1, 2009. We calculated the accuracy, overstaging rate, understaging rate, Kappa statistic, sensitivity, and specificity for abdominal ultrasound (AUS), computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) with respect to the gold

standard (pathology). We also compared the performance of CT by detector number and image type. A meta-analysis was performed.

Results For pre-operative T staging MRI scans had better performance accuracy than CT and AUS; CT scanners using ≥ 4 detectors and multi-planar reformatted (MPR) images had higher staging performances than scanners with < 4 detectors and axial images only. For pre-operative N staging PET had the lowest sensitivity, but the highest specificity among modalities; CT performance did not significantly differ by detector number or addition of MPR images. For pre-operative M staging performance did not significantly differ by modality, detector number, or MPR images.

Conclusions The agreement between pre-operative TNM staging by imaging scans and post-operative staging by pathology is not perfect and may affect treatment decisions. Operator dependence and heterogeneity of data may

Electronic supplementary material The online version of this article (doi:10.1007/s10120-011-0069-6) contains supplementary material, which is available to authorized users.

R. Seevaratnam · R. Cardoso · A. Mahar · N. Coburn
Sunnybrook Research Institute, Sunnybrook Health Sciences Centre, Toronto, Canada

C. McGregor
Department of Medical Imaging, Sunnybrook Health Sciences Centre, Toronto, Canada

L. Lourenco
Department of Surgery, Universidade Federal de Sao Paulo, Sao Paulo, Brazil

A. Mahar
Department of Community Health and Epidemiology, Queen's University, Kingston, Canada

R. Sutradhar
Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

R. Sutradhar · C. Law · L. Paszat · N. Coburn
Institute for Clinical Evaluative Sciences, Toronto, Canada

C. Law · N. Coburn (✉)
Division of Surgical Oncology, Odette Cancer Centre, Sunnybrook Health Sciences Centre, Suite T2-60, 2075 Bayview Ave, Toronto, ON M4N 3M5, Canada
e-mail: natalie.coburn@sunnybrook.ca

L. Paszat · N. Coburn
Department of Health Policy, Management and Evaluation, University of Toronto, Toronto, Canada

account for the variations in staging performance. Physicians should consider this discrepancy when creating their treatment plans.

Keywords Abdominal ultrasound (AUS) · Computed tomography (CT) · Magnetic resonance imaging (MRI) · Positron emission tomography (PET) · Gastric cancer

Introduction

Currently, resection is the only curative option available for patients with gastric cancer [1–4]. Accurate assessment of local tumor depth invasion (T), regional lymph node invasion (N), and distant metastases (M) is crucial to appropriate surgical and treatment planning [1, 2, 5]. Understaging of the disease may lead to positive resection margins or unnecessary laparotomy if metastases were not identified on pre-operative imaging. Overstaging a patient may lead to ineffective care if a potentially curative patient is incorrectly categorized as a palliative patient [5].

Available pre-operative staging modalities include abdominal ultrasound (AUS), computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET). Current National Comprehensive Cancer Network (NCCN) practice guidelines for gastric cancer [6] suggest using a variety of techniques as part of the workup, including CT of abdomen and pelvis, chest imaging, pelvic ultrasound, PET, PET-CT, esophagogastroduodenoscopy (EGD), and endoscopic ultrasound (EUS). However, the guidelines do not recommend specific modalities or workup pathways [6].

Despite the routine use of the above imaging modalities for pre-operative staging, each modality has limitations. AUS has difficulty in evaluating the wall of the gastric fundus and greater curvature, as well as lymphatic spread [7, 8]. It is also highly dependent on patient body habitus and the operator [9]. Traditional single detector scanners (S-CT) are limited by large section thickness, low image resolution, and slow scanning causing respiratory motion artifacts, and they are unable to provide multi-planar reformations [10, 11]. Multi-detector row CTs (MDCTs) have difficulty in detecting flat type lesions and have poor soft tissue contrast resolution [10, 11]. Nodal assessment is limited to size criteria, which does not allow diagnosis of microscopic nodal invasion or the exclusion of enlarged reactive nodes [10]. MRI scans have limitations including respiratory motion artifacts, long examination time, high costs, and lack of a standard gastric protocol [12, 13]. Assessment of nodal status by MRI is also limited to size criteria. Furthermore, MRI is limited in the amount of body coverage that can be achieved in a single exam, making it unsuitable for M staging [14]. 18-F-Fluoro-2-deoxyglucose

positron emission tomography (FDG-PET, simplified as PET throughout this paper) uses a semi-quantitative method, the standardized uptake value (SUV), to assess the uptake of FDG in a tumor [15]. However, SUVs are dependent on several factors, including time post FDG injection, tumor size, normoglycemia, and technical parameters [16, 17]. PET is also highly dependent upon the pathological subtype of the cancer, as mucinous tumors may give false-negative results [15].

The limitations of each technique have an effect on the ability of these modalities to accurately stage gastric cancer prior to surgery [1–4, 7, 8, 10–13, 15–17], yet patient operability and tumor resectability are heavily dependent on the quality of pre-operative imaging [5]. Therefore, the purpose of this review is to provide a detailed meta-analysis of the pre-operative TNM staging abilities of AUS, CT, MRI, and PET in patients with pathology-confirmed gastric cancer over the past decade.

Methods

Data sources

Electronic literature searches were conducted using Medline and Embase from January 1, 1998 to December 1, 2009, according to the search algorithm presented in Appendix A of electronic supplementary material. Search terms included: [exp stomach cancer/or (((gastric or stomach) adj1 cancer\$) or ((gastric or stomach) adj1 carcinoma) or ((gastric or stomach) adj1 adenocarcinoma) or ((gastric or stomach) adj1 neoplasm\$)).mp.] and [cancer staging/or diagnostic imaging/or exp computer assisted tomography/or computer assisted emission tomography/or exp positron emission tomography/or exp nuclear magnetic resonance imaging/or exp barium meal/] and [clinical trial/or controlled clinical trial/or exp comparative study/or meta analysis/or multicenter study/or exp practice guideline/or randomized controlled trial/] not [review or case report/] not [*gastrointestinal stromal tumor/or exp B cell lymphoma/and “marginal zone”.mp.]. A separate search of the Cochrane Central Register of Controlled Trials (1998–2009) was performed using the search term “gastric cancer”. Reference lists from review papers and relevant articles were also examined for additional studies that met our inclusion criteria.

Study selection and review process

To be eligible, studies had to meet the following criteria: (1) investigation of preoperative T, N, or M staging performance of AUS, CT, MRI, or PET in newly (not recurrent) diagnosed patients with histopathology-confirmed

gastric adenocarcinoma; (2) patients underwent surgery, and pre-operative staging was compared with post-operative pathological staging; (3) studies involved human patients with a minimum of 30 patients; and (4) studies were published in peer reviewed journals in English. Studies were excluded according to the following exclusion criteria: (1) studies that involved animals and/or ex vivo samples; (2) studies that involved patients with mixed cancer or studies investigating the diagnostic performance in other cancers with no separate analysis of gastric cancer subjects; (3) studies that did not provide sufficient information to determine pre-operative T, N, or M staging performance; and (4) review articles, meta-analyses, abstracts, conference proceedings, editorials/letters, and case reports. Studies that reported performance characteristics on more than one imaging technique were included only if the images from each technique were independently analyzed and the reviewers were blinded. All electronic search titles, selected abstracts, and full-text articles were

independently reviewed by a minimum of two reviewers (NC, LP, AM, RC, RS). Disagreements on study inclusion/exclusion were resolved with a consensus meeting.

Data extraction

A systematic approach to data extraction was used to produce a descriptive summary of participants, interventions, and study findings. The first reviewer (RS) independently extracted the data and a second reviewer (RC, CM) checked the data extraction. No attempt was made to contact authors for additional information. The TNM staging categories were extracted from corresponding publications. Staging classifications for individual studies can be found in Appendices 1–4 of the electronic supplementary material. Modality-specific staging definitions incorporated by the majority of studies are shown in Fig. 1. Both the Union International Contre le Cancer (UICC)/American Joint Committee on Cancer (AJCC) staging

T Stage	AUS Criteria ^a
T0	No evidence of alteration to any of the 5 gastric layers
T1	Layers 1 to 3 are interrupted or thickened while the 4th and 5th layers are normal sonographically
T2	4th layer is thickened while the 5th layer is intact sonographically
T3	Interruption or disappearance of all 5 layers of the gastric wall sonographically
T4	Tumor invading adjacent organs: has an indistinct border and is indistinguishable from involved organs sonographically
N Stage	AUS Criteria ^a
N0	No evidence of lymph node metastasis
N1	Metastasis in the perigastric node(s) closer than 3 cm to the primary lesion
N2	Metastasis in the perigastric node(s) distance of more than 3 cm to the primary lesion
N3	Metastasis in hepatoduodenal, peripancreatic, superior mesenteric, paraaortic, and middle colic nodes
M Stage	AUS Criteria ^a
M0	Distant metastasis absent
M1	Distant metastasis present

T Stage	MRI Criteria ^a
T0	No detectable primary lesion at the stomach wall
T1	Enhancing tumor does not penetrate the enhancing submucosal layer
T2	Clear continuous low signal intensity band or enhancing cancerous portion do not penetrate the low signal intensity band
T3	Interrupted low signal intensity band or enhancing cancerous portion penetrates the outer low signal intensity band
T4	Continuous extension of the cancerous portion to the adjacent organ with or without interruption of low signal intensity band
N Stage	MRI Criteria ^a
N0	No evidence of lymph node metastasis
N1	Metastasis in the perigastric node(s) closer than 3 cm to the primary lesion
N2	Metastasis in the perigastric node(s) distance of more than 3 cm to the primary lesion
N3	Metastasis in hepatoduodenal, peripancreatic, superior mesenteric, paraaortic, and middle colic nodes
M Stage	MRI Criteria ^a
M0	Distant metastasis absent
M1	Distant metastasis present

T Stage	CT Criteria ^b
T0	No evidence of alteration of the gastric wall with normal fat plane
T1	Neoplasm shows focal thickening of inner layer, is almost well enhanced, and has visible low-attenuation-strip outer layer of gastric wall and clear fat plane around tumor
T2	Neoplasm shows focal or diffuse thickening of gastric wall with transmural involvement, is almost well enhanced, and has smooth outer wall border and clear fat plane around tumor
T3	Transmural tumor with irregular or nodular outer border and/or perigastric fat infiltration
T4	Obliteration of fat plane between gastric tumor and adjacent organ or invasion of adjacent organ
N Stage	CT Criteria (described by the JGCA) ^b
N0	No evidence of lymph node metastasis
N1	metastasis in the perigastric nodes
N2	metastasis in the nodes along the celiac artery and its branches
N3	metastasis in the retropancreatic or para-aortic nodes
N stage	CT Criteria (described by AJCC/UICC) ^b
N0	No regional lymph node metastases
N1	Metastases in 1-6 regional lymph nodes
N2	Metastases in 7-15 regional lymph nodes
N3	Metastases in > 15 regional lymph nodes
M Stage	CT Criteria
M0	Distant metastasis absent
M1	Distant metastasis present

T Stage	PET Criteria ^d
T0	No increase in 18F-FDG uptake within the stomach
T+	Increased 18F-FDG uptake exceeding that of the adjacent normal gastric wall
N Stage	PET Criteria (described by the JGCA) ^d
N0	No evidence of lymph node metastasis
N1	Increased 18F-FDG uptake in the perigastric nodes
N2	Increased 18F-FDG uptake in the nodes along the celiac artery and its branches
N3	Increased 18F-FDG uptake in the retropancreatic or paraaortic nodes
N stage	PET Criteria (described by AJCC/UICC) ^d
N0	No regional lymph node metastases
N1	Increased 18F-FDG uptake in 1-6 regional lymph nodes
N2	Increased 18F-FDG uptake in 7-15 regional lymph nodes
N3	Increased 18F-FDG uptake in > 15 regional lymph nodes
M Stage	PET Criteria ^d
M0	Distant metastases absent
M1	Increased 18F-FDG uptake in distant metastases

Fig. 1 TNM staging criteria for gastric cancer by modality. ^aAdapted from [42, 43]. ^bAdapted from [28, 53]. ^cAdapted from [36]. ^dAdapted from [50, 56]. All articles adapted the above definitions or slight variations of these definitions when describing their respective modalities. AUS, abdominal ultrasound; CT, computed tomography;

JGCA, Japanese Gastric Cancer Association; AJCC, American Joint Committee on Cancer; UICC, Union International Contre le Cancer; MRI, magnetic resonance imaging; PET, positron emission tomography

classifications [18], and the Japanese Gastric Cancer Association (JGCA) 2nd English edition [19] classification system were used (Appendix B of the electronic supplementary material).

Data analysis

A range of definitions was found for the calculation of accuracy, sensitivity, and specificity. Therefore, the following performance characteristics were re-calculated from the original numbers provided in each included publication: detection rate, accuracy, overstaging rate, understaging rate, agreement/kappa statistic (κ), sensitivity, and specificity. Detection rate was defined as the ability to detect the presence of a tumor. Accuracy was defined as the ability to match the pre-operative stage of a given tumor with the post-operative pathology staging (i.e., T1 accuracy = [number correctly staged by pre-operative imaging technique as T1/number staged by pathology as T1] \times 100). Over- and understaging refer to when the tumor was incorrectly staged higher and lower compared to post-operative pathological staging, respectively. Overall calculations for accuracy, overstaging rate, and understaging rate were based on the average performance values for all cases (i.e., combined values for T1–T4; i.e., overall accuracy = [number of cases correctly staged/number of all cases] \times 100). Agreement between the pre-operative imaging technique and pathology was calculated using a 4 \times 4 table (corresponding to stages T1, T2, T3, and T4). A 5 \times 5 table was used when the pre-operative imaging technique did not detect the presence of a tumor (stage T0), while a 3 \times 3 table was used when two of the stages were combined (e.g., T1–T2). The following interpretation of κ was used: <0 = less than chance agreement; 0.01–0.20 = slight agreement; 0.21–0.40 = fair agreement; 0.41–0.60 = moderate agreement; 0.61–0.80 = substantial agreement; 0.81–0.99 = almost perfect agreement [20]. For pre-operative N staging, sensitivity and specificity of staging a lymph node as negative (N0) or positive (N+) was determined using a 2 \times 2 table (corresponding to N0 and N+). For pre-operative M staging, sensitivity and specificity of staging metastases as negative (M0) or positive (M1) was determined using a 2 \times 2 table (corresponding to M0 and M1). Overall calculations for sensitivity and specificity for N and M stage were based on the average values for all cases (i.e., N0–N+ and M0–M+). Statistical analyses were conducted using R version 2.10.1 statistical package (<http://cran.r-project.org/>). Meta-analysis (pooling of data) was calculated using the inverse variance method and the random effects estimate based on the DerSimonian–Laird method [21]. Only performance characteristics that were re-calculated were included in pooling analyses. Significance within and between imaging techniques was calculated by comparing pooled scores.

A Bonferroni correction was applied when multiple comparisons were made such that significance was reached when $P \leq \alpha/N$ (where $\alpha = 0.05$ and $N =$ number of comparisons/outcomes measured) [22].

Results

Literature search

A total of 5204 titles/abstracts were identified from the electronic searches and reference lists for preliminary review. After removal of duplicates and screening for relevant titles and abstracts, a total of 167 articles were submitted for a full-text review. A total of 40 articles [23–62] involving 3758 patients met our inclusion criteria and were included in this review (Fig. 2). We included 29 prospective studies and 11 retrospective studies.

Performance characteristics of pre-operative imaging studies

Overall TNM staging results for each technique (AUS, CT, MRI, and PET) are presented in Tables 1, 2, 3, and 4, respectively, with more detailed (stage/metastatic site specific) analyses found in the correspondingly numbered electronic Appendices 1–4 of the electronic supplementary material. For the evaluation of pre-operative diagnostic AUS (3 studies [37, 42, 43]), a total of 168 patients were assessed for T stage, 149 patients were assessed for N stage, and 101 patients were assessed for M stage pre-operatively by AUS and post-operatively by pathology. For the evaluation of pre-operative diagnostic CT (32 studies [23–35, 37–42, 44–49, 52–59]), a total of 2909 patients were assessed for T stage, 2646 patients were assessed for

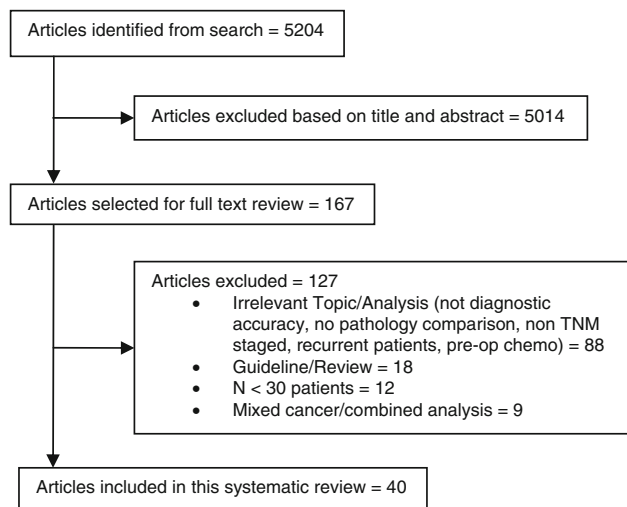


Fig. 2 Article selection flow

Table 1 Performance characteristics of abdominal ultrasound (AUS) studies

Study	Study type	Patients (N)	T1/T2 (%)	N0 (%)	M0 (%)	Transducer frequency	Stage	Overall accuracy (%)	Agreement (κ)	Overall overstaging (%)	Overall understaging (%)	Sensitivity	Specificity
Kayaalp [37]	P	118	-	-	39	NR	M	85.3	-	-	-	35.3	95.8
Lee [42]	P	43	60	58	-	3.5 MHz	T	55.8	0.398	13.9	30.3	44.4	92
Liao [43]	P	125	20	37	44	3.5–6.0 MHz	T	77.6	0.658	6.4	16	-	-
							N	72.6	-	13.2	14.2	77.6	64.1
							M	43.3	-	-	-	-	-

For stage breakdown performance characteristics and study details please see Appendix 1 of the electronic supplementary material
N number, *P* prospective, *NR* not reported

N stage, and 916 patients were assessed for M stage pre-operatively by CT and post-operatively by pathology. For the evaluation of pre-operative diagnostic MRI (3 studies [36, 49, 51]), a total of 109 patients were assessed for T stage, and 75 patients were assessed for N stage pre-operatively by MRI and post-operatively by pathology. For the evaluation of pre-operative diagnostic PET (9 studies [29, 44, 45, 50, 55, 56, 60–62]), a total of 422 patients were assessed for T stage, 420 patients were assessed for N stage, and 282 patients were assessed for M stage pre-operatively by PET and post-operatively by pathology.

Comparison of AUS, CT, MRI, and PET

The pooled TNM performance characteristics of all modalities are reported in Table 5. Overall, MRI had significantly better T staging performance compared to all CT scanners, as well as better T1 staging performance compared to AUS. Because PET cannot stage cancers by tumor depth, we calculated the primary tumor detection rate reported in all studies. This pooled value was $80.4 \pm 4.9\%$, with an overall detection rate ranging from 58.1 to 95.9% (Appendix 4.3 of the electronic supplementary material). The primary tumor detection rates for AUS, CT, and MRI ranged from 90.7–100, 61.1–100, and 97.8–100%, respectively (Appendices 1.3, 2.3, and 3.3 of the electronic supplementary material). For N staging, PET had the lowest sensitivity and the highest specificity. There was no superior modality for determining M stage.

Pre-operative TNM staging performance by detector number and use of multi-planar images

We compared the pooled TNM performance characteristics of CT scanners with <4 detectors [25, 26, 30–33, 35, 37, 40–42, 46, 49, 52, 57, 59] to those with ≥ 4 detectors [23, 24, 27, 28, 34, 38, 39, 47, 48, 53–55, 58], to determine whether the use of more detector rows to capture images translated into better pre-operative staging performances (Table 6). Overall, CT scanners with ≥ 4 detectors had significantly better T staging performances compared to CT scanners with <4 detectors. However, detector number did not significantly affect N or M staging performances.

We compared the pooled TNM performance characteristics of CT scanners using traditional single plane axial images [25, 28–35, 37, 38, 40–42, 46, 49, 52, 57–59] with scanners using multi-planar reformatted (MPR) images [23, 24, 27, 28, 34, 38, 39, 44, 47, 48, 53–56, 58], to determine whether the addition of multiple image planes translated into better pre-operative staging performances (Table 6). Overall, CT scanners using MPR images had significantly better T staging performances compared to

Table 2 Performance characteristics of computed tomography (CT) studies

Study	Study type	Patients (N)	T1/T2 (%)	N0 (%)	M0 (%)	DT (N)	Image type	Stage	Overall accuracy (%)	Agreement (κ)	Overall overstaging (%)	Overall understaging (%)	Overall sensitivity	Overall specificity
Ahn [23]	R	434	88 ^a	89	-	16	3D, A, MPR	T	86.4	0.35	6.7	6.9	-	-
Bhandari [24]	P	48	69	58	-	4	3D, A, MPR, VE	T	83.6	-	7.4	9	17	91.7
Blackshaw [25]	P	100	15	-	54	1	A	N	83.3	0.687	16.7	0	-	-
Cereceda Perez [26]	P	46	13	NR	80	1	MLR, VE	M	75	-	10.4	14.6	80	85.7
Chen [27]	R	64	56	47	-	16	3D, A, MPR, VE	T	60	0.336	20	20	-	-
Chen [28]	P	55	47	33	84	16	3D, A, MPR, VE	M	40.8	-	-	-	46.2	100
Chen [29]	P	68	-	19	82	1, 4	A	T	56.5	NR	19.6	23.9	-	-
D'Elia [30]	P	107	47	33	84	1	A	N	56	-	4	18	62.5**	92.9**
Dux [31]	P	112	61	32	64	1	A	M	87	-	-	-	88.9**	86.5**
Gamon-Giner [32]	P	50	26	18	88	1	A	T	73.4	0.627	11	15.6	-	-
Habermann [33]	P	51	57	38	-	1	A	N	68.8	-	15.6	15.6	88.2	80
Hundt [57]	P	40	50	5	-	1	A	T	MPR = 89	MPR = 0.8	MPR = 5.5	MPR = 5.5	-	-
								N	A = 72.7	A = 0.589	A = 5.5	A = 21.8	MPR = 86.5	MPR = 77.8
								M	MPR = 78.2	-	MPR = 12.7	MPR = 9.1	A = 91.9	A = 72.2
								N	A = 70.9	-	A = 10.9	A = 18.2	53.6	100
								M	98	-	-	NR	LN = 78.2**	LN = 61.5**
								N	81.8	-	NR	NR	DN = 75**	DN = 90.6**
								M	75	-	-	-	60	96.1
								T	77.6	0.677	14	8.4	-	-
								N	71.1	-	16.8	12.1	97.2	65.7
								M	98.1	-	-	-	57.9	99.5
								T	51.1	0.68	25.5	23.4	-	-
								N	50.6	-	12.9	36.5	63.8	74.1
								M	78.7	-	-	-	71.8	82.6
								T	56	0.321	16	28	-	-
								N	82	-	4	14	82.9	77.8
								M	84	-	-	-	50	88.6
								T	76.5	0.57	13.7	9.8	-	-
								N	70	-	14	16	74.2	84.2
								T	79.5	0.657	7.7	12.8	-	-
								N	69.2	-	7.7	23.1	94.6	100

Table 2 continued

Study	Study type	Patients (N)	T1/T2 (%)	N0 (%)	M0 (%)	DT (N)	Image type	Stage	Overall accuracy (%)	Agreement (κ)	Overall overstaging (%)	Overall understaging (%)	Overall sensitivity	Overall specificity
Hur [34]	P	70	63	46	-	16	A, MPR	T	MPR = 77.1 A = 67.1	NR	MPR = 14.3 A = 18.6	MPR = 8.6 A = 14.3	-	-
Kadowaki [35]	R	59	71	-	-	1	A	T	MPR = 59 A = 53.7	-	MPR = 17 A = 15.9	MPR = 24 A = 30.4	NR	NR
Kayaalp [37]	P	46	-	-	100	1	A	M	66.1 85.8	-	20.3	13.6	38	95.3
Kim [38]	P	106	73	57	9	4	A, MPR, VE	T	V = 82.1 A = 67	V = 0.734 A = 0.542	V = 9.4 A = 10.4	V = 8.5 A = 22.6	-	-
								N	V = 64.2 A = 62.3	-	V = 24.5 A = 23.6	V = 11.3 A = 14.1	V = 71.7 A = 78.3	V = 63.3 A = 61.7
								M	V = 85.7 A = 85.7	-	-	-	V = 85.7 A = 100	V = 100 A = 100
Kim [58]	P	174	-	76 ^c	-	40	A, MPR	N	EGC: MPR = 81.4 A = 75.5 AGC: MPR = 52.8 A = 41.7	-	EGC: MPR = 4.9 A = 6.9 AGC: MPR = 18.1 A = 18.1	EGC: MPR = 13.7 A = 17.6 AGC: MPR = 29.1 A = 40.2	EGC: MPR = 75 A = 50 AGC: MPR = 83.3 A = 71.4	EGC: MPR = 96.2 A = 91 AGC: MPR = 63.3 A = 60
Kumano [39]	R	41	76	-	4	4	A, MPR	T	82.9	0.694	10.9	6.2	-	-
Lee [40]	P	40	40	35	-	1	3D, A	T	70	0.511	2.5	27.5	-	-
Lee [41]	P	180	57	47	-	1	A	T	72.5 47.8	-	5 23.3	22.5 28.9	80	85.7
Lee [42]	P	43	60	58	-	1	A	N	52.2 58.2	-	16.1 13.9	31.7 27.9	-	-
Lim [44]	R	112	-	-	86	1, 4, 16	A, MPR	M	55.8 89.3	-	18.6	25.6	55.6 76.5	72 91.6
Polkowski [46]	P	88	23	23	-	2	A	T	44.3 46.7	0.134	28.4	27.3	-	-
Shimizu [47]	P	60	77	-	4	4	A, MPR	T	85	0.756	8.3	6.7	82.6	50
Shinohara [48]	P	112	95 ^b	23	81	4	A, MPR	T	95.5 58.9	0.592	2.7 13.4	1.8 27.7	-	69.2 100
Sohn [49]	P	30	37	NR	-	1	A	M	57.1 66.7	0.488	10	23.3	-	-
Takao [59]	P	108	72	-	-	1	A	T	58.6 49.1	-	6.9	34.5	NR	NR
Wei [52]	R	50	48	-	-	1	A	T	82	0.407	5.5	45.4	-	-
								T	82	0.73	10	8	-	-

Table 2 continued

Study	Study type	Patients (N)	T1/T2 (%)	N0 (%)	M0 (%)	DT (N)	Image type	Stage	Overall accuracy (%)	Agreement (κ)	Overall overstaging (%)	Overall understaging (%)	Overall sensitivity	Overall specificity
Yan [53]	P	790	34	41	-	4	A, MPR	T	76.5	0.57	13.7	9.8	-	-
Yang [54]	P	44	61	57	-	64	A, MPR	T	75.3	-	12.2	12.5	86.3	76.2
Yang [55]	R	68	-	44	-	16	A, MPR	N	84.1	0.763	9.1	6.8	-	-
Yun [56]	R	81	-	35	-	1, 4, 16	A, MPR	N	79.5	-	9.1	11.4	84.2	84
								N	54.4	-	7.4	38.2	60.5	83.3
								N	77.8	-	NR	NR	50.8**	91.2**

For stage breakdown performance characteristics and study details please see Appendix 2 of the electronic supplementary material
N number, *P* prospective, *R* retrospective, *DT* detectors, *3D* 3-dimensional, *A* axial, *MPR* multi-planar reformatted, *VE* virtual endoscopy, *NR* not reported, *MZR* multi-level reconstruction, *LN* local lymph node group, *DN* distant lymph node group, *EGC* early gastric cancer, *AGC* advanced gastric cancer
 ** Values from calculations published in original manuscript

^a % for T1 patients only as reported in publication

^b % for T1-T3 patients as reported in publication

^c EGC

^d AGC

Table 3 Performance characteristics of magnetic resonance imaging (MRI) studies

Study	Study type	Patients (N)	T1/T2 (%)	N0 (%)	IV contrast (Gd-DTPA)	Stage	Overall accuracy (%)	Agreement (κ)	Overall overstaging (%)	Overall understaging (%)	Sensitivity	Specificity
Kang [36]	P	46	37	26	0.1 mmol/kg	T	82.6	0.751	2.2	15.2	-	-
Sohn [49]	P	30	37	45	15-20 mg at 1 ml/s	T	52.2	-	17.4	30.4	85.3	75
Wang [51]	P	33	18	-	0.1 mmol/kg	T	73.3	0.607	6.7	20	-	-
						N	55.2	-	10.3	34.5	NR	NR
						T	87.8	0.776	6.1	6.1	-	-

For stage breakdown performance characteristics and study details please see Appendix 3 of the electronic supplementary material

N number, *IV* intravenous, *P* prospective, *Gd-DTPA* gadopentetate dimeglumine, *NR* not reported

Table 4 Performance characteristics of positron emission tomography (PET) studies

Study	Study type	Patients (N)	T1/T2 (%)	N0 (%)	M0 (%)	IV contrast (18F-FDG)	Overall detection rate (%)	Stage	Overall mean SUV	Overall accuracy (%)	Overall overstaging (%)	Overall understaging (%)	Sensitivity	Specificity
Chen [29]	P	68	12	21	97	370–555 MBq	94.1	T	7	–	–	–	–	–
								N	6.8	79.2	NR	NR	LN = 56.4** DN = 87.5**	LN = 92.3** DN = 96.2**
Kim [60]	P	73	33	12	92	555 MBq	95.9	M	6.5	100	–	–	100	100
								T	7.68	–	–	–	–	–
								N	4.52	NR	NR	NR	40**	95.2**
Lim [44]	R	112	NR	NR	86	370–555 MBq	NR	M	NR	NR	–	–	16.7**	98.9**
Mochiki [45]	P	85	67	0	–	275–370 MBq	75.3	M	NR	89.3**	–	–	35.3**	98.9**
								T	4.25	–	–	–	–	–
Mukai [61]	R	75	57	53	–	4.44 MBq/kg	58.1	N	4.68	NR	NR	NR	23.3**	100**
Stahl [62]	R	40	0	–	–	300 MBq	60	N	NR	17.7	NR	NR	34.5**	97**
								T	4.68	–	–	–	–	–
Tian [50]	P	30	NR	19	70	148–185 MBq	NR	N	4.95	73.3	0	26.7	52.9	100
Yang ^a [55]	R	78	NR	46	–	200 MBq	NR	M	6.5	100	–	–	100	100
								N	NR	55.1	1.3	43.6	30.9	97.2
Yun [56]	R	91	21	0	–	370 or 518 MBq	87.7	T	5.4	–	–	–	–	–
								N	NR	74.1	NR	NR	39.5**	29.1**

For stage breakdown performance characteristics and study details please see Appendix 4 of the electronic supplementary material

N number, P prospective, NR not reported, LN local lymph node, DN distant lymph node

** Values from calculations published in original manuscript

^a Values reported for PET-CT techniques combined (not PET alone)

Table 5 Comparison of performance characteristics by imaging technique

	AUS	CT	MRI	PET
T staging				
κ	0.54 ± 0.13	0.55 ± 0.04	0.73 ± 0.06^a	–
Overall accuracy (%)	67.8 ± 10.8	71.5 ± 2.7	82.9 ± 3.7^a	–
T1 accuracy (%)	45.4 ± 10.4	63.0 ± 5.2	86.3 ± 5.3^{a,b}	–
T2 accuracy (%)	84.4 ± 7.3	72.9 ± 3.6	76.7 ± 6.3	–
T3 accuracy (%)	73.9 ± 15.4	75.3 ± 2.6	86.8 ± 3.3^a	–
T4 accuracy (%)	52.7 ± 31.2	74.9 ± 3.9	80.2 ± 12.8	–
N staging				
Overall accuracy (%)	68.1 ± 5.8	66.1 ± 2.1	53.4 ± 5.9	60.0 ± 10.8
Sensitivity (%)	63.0 ± 16.5	77.2 ± 2.6^c	85.3 ± 4.7^{#,d}	40.3 ± 10.9
Specificity (%)	78.8 ± 13.9	78.3 ± 2.5	75.0 ± 9.3 [#]	97.7 ± 1.3^e
M staging				
Overall accuracy (%)	64.7 ± 21.0	81.2 ± 3.4	–	88.2 ± 5.8

Data presented as pooled means ± standard error of all studies within a given category using the inverse variance method. Studies that did not report raw data for re-calculation were excluded from the pooling analysis. Pooled T and M staging performance characteristics were not available for PET and MRI, respectively. Bonferroni correction: $P \leq 0.025$ (2 comparisons) and $P \leq 0.016$ (3 comparisons). Pooled detection rate for PET was $80.4 \pm 4.9\%$. Values in bold indicate significant differences

[#] Only one study reported values (i.e., not pooled value)

^a MRI significantly higher ($P \leq 0.014$) versus CT

^b MRI significantly higher ($P \leq 0.0004$) versus AUS

^c CT significantly higher ($P \leq 0.001$) versus PET

^d MRI significantly higher ($P \leq 0.0001$) versus PET

^e PET significantly higher versus CT ($P < 0.0001$) and MRI ($P = 0.016$)

Table 6 Comparison of computed tomography performance characteristics by detector number and MPR images

	<4 detectors	≥4 detectors	Axial images	MPR images
T staging				
κ	0.45 ± 0.05	0.65 ± 0.03^a	0.46 ± 0.04	0.67 ± 0.04^c
Overall accuracy (%)	62.8 ± 3.6	80.4 ± 2.7^a	65.2 ± 3.3	81.9 ± 3.1^c
T1 accuracy (%)	47.5 ± 11.8	75.2 ± 5.2^a	52.9 ± 9.9	76.4 ± 5.7^c
T2 accuracy (%)	65.9 ± 5.5	80.0 ± 5.0 ^b	70.4 ± 4.8	77.7 ± 6.3
T3 accuracy (%)	69.3 ± 3.8	84.5 ± 2.4^a	71.6 ± 3.4	85.3 ± 2.7^c
T4 accuracy (%)	71.8 ± 6.1	78.8 ± 6.2	69.1 ± 6.5	83.5 ± 5.0 ^d
N staging				
Overall accuracy (%)	62.1 ± 3.5	67.1 ± 2.6	63.2 ± 3.0	70.5 ± 2.6 ^d
Sensitivity (%)	79.1 ± 4.2	75.8 ± 3.4	77.1 ± 3.7	77.0 ± 4.1
Specificity (%)	76.4 ± 5.6	78.8 ± 2.9	75.9 ± 4.4	80.1 ± 3.3
M staging				
Overall accuracy (%)	79.4 ± 7.5	82.2 ± 7.3	78.6 ± 6.5	83.2 ± 6.6

Data presented as pooled means ± standard error of all studies within a given category using the inverse variance method. Studies that did not report raw data for re-calculation and studies that did not specify slice number or axial and MPR (multi-planar reformatted) image type were excluded from the pooling analyses. Values in bold indicate significant differences

^a ≥4 Slices significantly higher ($P \leq 0.03$) versus <4 slices

^b Trend with ≥ 4 slices higher ($P = 0.056$) versus <4 slices

^c MPR images significantly higher ($P \leq 0.04$) versus axial images

^d Trend with MPR images higher versus axial images for T4 ($P = 0.079$) and N staging overall accuracy ($P = 0.068$)

axial images. However, additional MPR images did not significantly affect N or M staging performances.

Discussion

Accurate assessment of pre-operative TNM staging in gastric cancer is crucial for determining appropriate treatment strategies, especially for planning surgery, which remains the foundation for cure [1–4]. We reviewed a total of 40 studies (3758 patients): 3 AUS studies (168 patients) [37, 42, 43], 32 CT studies (2909 patients) [23–35, 37–42, 44–49, 52–59], 3 MRI studies (109 patients) [36, 49, 51], and 9 PET studies (422 patients) [29, 44, 45, 50, 55, 56, 60–62] on their pre-operative TNM staging performance values over the past decade (Tables 1–4/Appendices 1–4 of the electronic supplementary material).

TNM staging classifications

This review includes studies published over a span of 10 years and as such many TNM staging classifications are utilized (Appendix B of the electronic supplementary material). There are no differences between the 3rd and 4th editions of the UICC/AJCC system, which were incorporated by 25% of the included studies. Although the 6th edition divides T2 into T2a and T2b, the studies included in this review did not incorporate this breakdown; for the purposes of this review, the 5th and 6th editions are considered the same (37.5% of the included studies) [18]. The main difference between the 3rd/4th and 5th/6th UICC/AJCC editions is the classification of N stage. The 3rd/4th editions did not have an N3 stage, and the N1 and N2 stages were defined according to the distance of the perigastric regional lymph nodes from the edge of the primary tumor [18]. The 5th/6th editions defined N1, N2, and N3 stages according to the total number of lymph node metastases present. Additionally, the 5th/6th editions considered metastases to the hepatoduodenal nodes as regional lymph nodes, whereas the 3rd/4th editions considered them as distant metastases (M1 disease) [18]. The 2nd English edition of the JGCA classification system was utilized in 30% of the studies [19]. The main differences between the JGCA and UICC/AJCC systems are the classifications for N and M stage. The JGCA defines N1, N2, and N3 stages according to the lymph node groups with respect to the location of the primary tumor [19]. In general, Group 1 nodes refer to the perigastric nodes, Group 2 nodes refer to the nodes along the celiac artery and its branches, and Group 3 nodes refer to the retropancreatic or paraaortic nodes, whereas in the UICC/AJCC classification, retropancreatic and paraaortic nodes are classified as distant metastases (M1 disease). Furthermore, the 2nd English edition of the JGCA system does not consider peritoneal, liver, and

cytological metastases as M1 disease (although the presence of these indicates stage IV disease), whereas the UICC/AJCC system does [18, 19]. Finally, 7.5% of the included studies used other staging classifications, such as those adopted by the World Health Organization (WHO), as well as those created by the former JGCA (the Japanese Research Society for Gastric Cancer; JRS GC) in 1993 and 1995.

Despite the incorporation of various classification systems, this is not a limitation in our meta-analysis. The T stage breakdown across all editions is the same, because the T2a and T2b definitions were not incorporated; thus, the pooling of data and comparison between studies was not affected. Due to the various N stage classifications, our meta-analysis only compared the ability to identify N0 versus N+ disease, as these definitions are consistent across all systems, thus making it possible to compare studies. With respect to M stage, our meta-analysis utilized the UICC/AJCC and not the JGCA definitions, and thus considered peritoneal, liver, and cytological metastases as M1 disease. Re-classification as M1 was possible for the included studies utilizing the JGCA definitions because the presence of peritoneal, liver, and cytological metastases was mentioned within the publications.

Evaluation of T staging

The value of AUS in pre-operative T staging remains unclear. We did not find any significant differences between AUS and the other imaging modalities, except for poor T1 staging performance compared to MRI (Table 5). The lack of significance is most likely attributable to the large standard error and limited published studies. We included only 2 studies that reported pre-operative T staging values (Appendix 1 of the electronic supplementary material), one of which had fair agreement (κ , 95% confidence interval [CI]: 0.40, 0.20–0.60) and another with substantial agreement (0.66, 0.55–0.77). This variation may be explained by the difficulty in staging tumors found in the gastric fundus and greater curvature [7, 8, 42, 43], as well as the highly subjective nature of AUS staging and thus its strong operator dependence [42, 43].

The T staging performance characteristics of CT scanners are moderate, with a pooled κ of 0.55, an overall accuracy of 71.5%, and stage-specific accuracies ranging from 63 to 75% (Table 5). However, when taking detector number and MPR images into consideration, the performance value of CT is improved (Table 6). Specifically, the use of ≥ 4 detector scanners results in a substantial increase in agreement with pathology ($\kappa = 0.65$), an overall accuracy of 80%, and stage-specific accuracies ranging from 75 to 84.5%. These results are supported by other studies that have shown similar improvements in T staging with increased detector number [10, 11, 13, 63]. Therefore, we

recommend that pre-operative T staging of gastric cancer be performed on MDCT scanners with ≥ 4 detectors. If determination of organ invasion is necessary, a higher-capacity scanner may give more accurate results (T1: 75.2 vs. 47.5% and T3: 84.5 vs. 69.3% for ≥ 4 detector and < 4 detector scans, respectively, Table 6). Accurate staging of T1 versus T2 is important for endoscopists considering endoscopic mucosal resection (EMR) and may be aided by using EUS [64, 65], whereas accurate staging of T3 versus T4 is important as the surgeon would need to plan a multi-visceral resection. The use of MPR images significantly improved T staging performance compared to axial images alone, resulting in a substantial agreement with pathology ($\kappa = 0.67$), an overall accuracy of 82%, and stage-specific accuracies ranging from 76 to 85% (Table 6). These results are supported by other studies that have shown similar improvements in T staging with multiple image planes [10, 11, 13, 63]. Therefore, we recommend that MPR images be included in the protocol for pre-operative T staging of gastric cancer if determination of T stage is critical.

Our results show that MRI had the best overall performance characteristics for T staging compared to other staging modalities, with a substantial agreement with pathology ($\kappa = 0.73$), an overall accuracy of 83%, and a stage-specific accuracy ranging from 77 to 87% (Table 5). However, it is important to note that only 3 MRI studies examining 109 patients were included in this review. Therefore, while the pre-operative T staging ability of MRI is highly accurate, a publication bias may be present, as all 3 studies reported excellent results (compared to the literature found on CT scans which included publication of poor results), which may have caused an overestimation of its performance abilities. Furthermore, current MRI protocols are breath-hold-dependent [12, 13, 36, 49, 51]; as such, it is possible that the patient cohorts included in these studies were better able to comply than the gastric cancer population as a whole.

Despite the inability to stage gastric cancer by tumor depth, PET has a pooled primary tumor detection rate of 80%, which suggests a good overall ability for identifying a gastric cancer if one exists. Not surprisingly, PET has a higher capacity to detect advanced gastric tumors (83–100%) compared to early gastric tumors (26–63%; see Appendix 4 of the electronic supplementary material). However, the ability of PET to detect various pathological tumors varies greatly with type: intestinal type (65.5–83%), non-intestinal type (41–79%), poorly differentiated adenocarcinomas (61.5–79%), and signet ring cell carcinoma (0–78%; see Appendix 4 of the electronic supplementary material).

It is important to mention that the overall accuracy of T staging for a given study (and for the pooled population) is dependent on the distribution of T stage within the

evaluated patient population. Typically, T1 and T2 accuracies are generally lower than those for T3 and T4 because of the inability to discriminate depths of invasion in early cancers. The relationship between T1 versus T2 tumors and T3 versus T4 tumors, however, is more complicated and sensitive to operator performance and imaging modality. In our meta-analysis, a significant difference between T-staged groups was not found (data not shown). Nonetheless, with the exception of MRI, a visible trend was found for higher T3 compared to T1 accuracies across modalities (Table 5). However, exceptions to this trend have been documented. For example, Table 2 shows that Ahn et al. [23] had a high overall accuracy of 86.4%, with 88% of the patients staged as T1, while Blackshaw et al. [25] had a low overall accuracy of 60%, with 85% of the patients staged as T3/T4. Consistent with the concept of being able to differentiate early versus advanced tumors, the distribution of the patient population can also have an effect on the sensitivity and specificity of identifying lymph nodes. In a patient population with a greater number of advanced tumors, it is likely that there will be a higher sensitivity and specificity for identifying lymph node involvement compared to a population with a greater number of early tumors, due to a higher pre-test probability of nodal involvement.

Evaluation of N staging

The ability to stage lymph node (LN) status pre-operatively in gastric cancer patients remains poor. Our results show that imaging modalities range in overall accuracy from 53% (MRI) to 66% (CT), in sensitivity from 40% (PET) to 85% (MRI), and in specificity from 75% (MRI) to 98% (PET), with no significant differences between modalities (Table 5). The specificities for all modalities were higher than their respective sensitivities. Among CT scanners, neither detector number nor MPR images significantly improved N staging (Table 6). The 85% sensitivity reported for MRI is from only one study, and thus it cannot be stated that MRI is clearly superior to other modalities. PET had the worst sensitivity (40%) of differentiating N0 and N+ nodes, but the best specificity (98%), suggesting it may be used to clarify true positive patients. These results confirm the analysis of another review that showed neither AUS, MDCT, conventional MRI, nor PET could reliably confirm or exclude the presence of LN metastasis [66]. Tumor-positive LNs are not always enlarged, and certain enlarged LNs are not always tumor-positive but instead are enlarged due to inflammation, both of these possibilities make N staging extremely difficult [15, 66]. Moreover, there are varying LN size criteria (ranging from > 6 mm to > 1 cm) required for LN detection [10]. We found that the majority (68%) of the studies incorporated a definition of

≥ 8 mm for LN involvement, although this criterion was applied to the short axis diameter in some cases and the long axis diameter in other cases (Appendix 2.4 of the electronic supplementary material). These size-dependent diagnostic criteria for AUS, CT, and MRI may also contribute to the lower specificity found among these modalities compared to PET, which utilizes a metabolic diagnostic criterion. However, the mean SUV noted for N staging can also vary, with overall values ranging from 4.5 to 6.8 (Table 4), and mean SUVs overlapping between N stage categories (N0: 3.5–6.0; N1: 2.7–7.5; N2: 4.5–9.0; N3: 6.2–8.7; Appendix 4.4 of the electronic supplementary material). These inaccuracies in true nodal status make pre-operative determination of disease spread difficult, and must be taken into account in reports of pre-operative staging for neoadjuvant and peri-operative treatments, as well as in the selection of patients for EMR in early gastric cancer. However, the progress made in the field of molecular biology, where studies have successfully documented the ability to use specific radio-labeled probes to tag and identify specific tumor antigens and/or receptors [67], will undoubtedly contribute to the advancement of pre-operative staging in gastric cancer, which should lead to more effective staging strategies in the future.

Evaluation of M staging

Currently, pre-operative M staging of gastric cancer can be best assessed by PET and ≥ 4 detector CT (overall accuracies of 88 and 82%, respectively; Tables 5 and 6). However, only 3 PET studies, compared to 11 CT studies, reported M staging accuracies. The value of AUS in M staging remains unclear. It had a pooled overall accuracy of 65%, but only 2 studies evaluated its potential (Table 1), resulting in high variation. The value of MRI in M staging was not assessed in any studies evaluated. In practice, MRI is not suitable for screening for metastases because of the limited area of the body that can be scanned in a single session [14]. However, it is often used to characterize non-specific liver lesions found by CT [68]. A limitation of this review is that pre-operative staging studies were included only if patients had a post-operative pathology report for comparison. Patients who were not offered curative resection on the basis of metastases found on pre-operative imaging were excluded. Therefore, the false-positive rate for metastatic disease may indeed be higher for all imaging techniques.

Overall

Despite the reasonable T and M staging abilities of CT, MRI, and PET, all are far from perfect. The importance of accurate pre-operative TNM staging has been demonstrated by

studies that show pre-operative staging frequently differs from post-operative assessments. Schwarz [5] found that post-operative assessment (based on intra-operative findings and pathology) differed from pre-operative staging in 29% of patients. In 45% of the cases, the changes in curative intent could be traced to uncertainty of diagnosis or disease extent [5]. Furthermore, 45.5% of patients with pre-operative stage assignment were ultimately re-classified into a different pathologic stage category post-operatively, and patients undergoing a curative-intent procedure were re-staged 50.4% of the time intra-operatively [5]. These high re-staging rates support the use of diagnostic laparoscopy (DL) to clarify pre-operative intent. Our review on the use of DL in gastric cancer found that DL changed management in 10–60% of cases [69].

It may be possible to increase the accuracy of pre-operative assessments by using combined staging modalities. van Vliet et al. found that the performance of CT alone was not sensitive enough for the detection of distant metastases, whereas the performance of AUS, neck US, and chest X-ray, in combination with CT resulted in higher accuracies in patients with esophageal or gastric cardia cancer [70]. Chen et al. [29] reported that the combined use of PET and CT was more accurate for pre-operative N and M staging than either modality alone; however, the combined pre-operative staging accuracy was still low, 66%. Therefore, further research is required to determine whether pre-operative TNM staging is improved by using combined and/or multiple imaging techniques.

Finally, the performance characteristics of a staging modality are determined by both the experience of the investigator and the quality of protocols, as well as by the equipment. Blackshaw et al. [63] found that pre-operative TNM staging by CT improved significantly with radiologist experience, with lower agreements in the first 75 patients compared to the last 25 patients staged. In this examination of the learning curve, the authors reported a twofold improvement in tumor detection and a sevenfold improvement in suspicious LN detection [63]. Variations in CT scanning protocols and equipment have been reported by Callaway and Bailey [71]. These authors surveyed 5 cancer networks (21 hospitals) covered by the South West Cancer Intelligence Service of the United Kingdom [71]. They found variation in the following: MDCT capabilities, gastric cancer patient volume, number of radiologists in each institution, radiologist specialty, CT scanning protocol, and image type used to evaluate patients [71]. Variations in scanning protocols included the use of various positive (gastrograffin vs. barium) and negative (water vs. milk) oral contrasts, execution of a pre-contrast scan, timing of scans (arterial phase vs. portal phase vs. both), and scan location (chest vs. abdomen vs. pelvis) [71]. Given these results, it is clear that pre-operative TNM

staging studies contain heterogeneous data, which may explain the high variation in performance characteristics reported by the studies included in our review. Importantly, the results reported in journal articles are likely better than those achieved on average. The publication of current and clear guidelines/protocols for routinely used imaging techniques is advocated.

Conclusion

The agreement between pre-operative TNM staging by radiology imaging and post-operative staging by pathology is far from perfect. *For pre-operative T staging* the performance characteristics of AUS and CT were not significantly different; however, MRI had a better performance, although in a limited number of patients. Among CT scanners, those using ≥ 4 detectors and MPR images performed better than scanners with < 4 detectors and axial images only. *For pre-operative N staging* overall accuracy was not significantly different across modalities; however, PET had the worst sensitivity yet highest specificity among modalities. CT performance did not significantly differ by detector number or addition of MPR images. *For pre-operative M staging* performance did not significantly differ by modality, detector number, or addition of MPR images. However, the lack of significance was most likely due to large standard errors. Operator dependence and heterogeneity of data may account for the variations in staging performance. Physicians should consider the implications of staging inaccuracy, and may want to use multiple imaging modalities and/or DL to confirm the specifics of a tumor prior to developing treatment strategies for gastric cancer patients.

Acknowledgments This research is funded by the Canadian Cancer Society (Grant #019325). Both Dr. Coburn (Career Scientist Award) and Dr. Paszat (Clinical Scientist Support) have received funding through the Ontario Ministry of Health and Long-Term Care. Dr. Law is supported by the Hanna Family Chair in Surgical Oncology.

References

- Ly QP, Sasson AR. Modern surgical considerations for gastric cancer. *J Natl Compr Cancer Netw*. 2008;6(9):885–94.
- Kovoor PA, Hwang J. Treatment of resectable gastric cancer: current standards of care. *Expert Rev Anticancer Ther*. 2009;9(1):135–42.
- Hartgrink HH, Jansen EP, van Grieken NC, van de Velde CJ. Gastric cancer. *Lancet*. 2009;374(9688):477–90.
- Catalano V, Labianca R, Beretta GD, Gatta G, de Braud F, Van Cutsem E. Gastric cancer. *Crit Rev Oncol Hematol*. 2009;71(2):127–64.
- Schwarz RE. Factors influencing change of preoperative treatment intent in a gastrointestinal cancer practice. *World J Surg Oncol*. 2007;5:32.
- National Comprehensive Cancer Network N. NCCN Clinical Practice Guidelines in Oncology—Gastric Cancer—V.2.2010. National Comprehensive Cancer Network, Inc.; 2010.
- Lim JH, Ko YT, Lee DH. Transabdominal US staging of gastric cancer. *Abdom Imaging*. 1994;19(6):527–31.
- Suk KT, Lim DW, Kim MY, Park DH, Kim KH, Kim JM, et al. Thickening of the gastric wall on transabdominal sonography: a sign of gastric cancer. *J Clin Ultrasound*. 2008;36(8):462–6.
- Solbiati L, Tonolini M, Cova L, Goldberg SN. The role of contrast-enhanced ultrasound in the detection of focal liver lesions. *Eur Radiol*. 2001;11(Suppl 3):E15–26.
- Chen CY, Wu DC, Kang WY, Hsu JS. Staging of gastric cancer with 16-channel MDCT. *Abdom Imaging*. 2006;31(5):514–20.
- Kim AY, Kim HJ, Ha HK. Gastric cancer by multidetector row CT: preoperative staging. *Abdom Imaging*. 2005;30(4):465–72.
- Tunaci M. Carcinoma of stomach and duodenum: radiologic diagnosis and staging. *Eur J Radiol*. 2002;42(3):181–92.
- Kwee RM, Kwee TC. Imaging in local staging of gastric cancer: a systematic review. *J Clin Oncol*. 2007;25(15):2107–16.
- Koh DM, Collins DJ. Diffusion-weighted MRI in the body: applications and challenges in oncology. *AJR Am J Roentgenol*. 2007;188(6):1622–35.
- 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(5):449–55.
- Thie JA. Understanding the standardized uptake value, its methods, and implications for usage. *J Nucl Med*. 2004;45(9):1431–4.
- Boellaard R, Krak NC, Hoekstra OS, Lammertsma AA. Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. *J Nucl Med*. 2004;45(9):1519–27.
- UICC/AJCC. American Joint Committee on Cancer, Cancer Staging Manual. 1988–2002. <http://www.cancerstaging.org/products/pasteditions.html>.
- Japanese Gastric Cancer Association (JGCA). Japanese classification of gastric carcinoma, 2nd English edition. *Gastric Cancer*. 1998;1:10–24.
- Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med*. 2005;37(5):360–3.
- Copper H, Hedges L. *The Handbook of Research Synthesis*. Newbury Park: Russell Sage Foundation; 1994.
- Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ*. 1995;310(6973):170.
- 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(1):20–7.
- Bhandari S, Shim CS, Kim JH, Jung IS, Cho JY, Lee JS, et al. Usefulness of three-dimensional, multidetector row CT (virtual gastroscopy and multiplanar reconstruction) in the evaluation of gastric cancer: a comparison with conventional endoscopy, EUS, and histopathology. *Gastrointest Endosc*. 2004;59(6):619–26.
- Blackshaw GR, Barry JD, Edwards P, Allison MC, Thomas GV, Lewis WG. Laparoscopy significantly improves the perceived preoperative stage of gastric cancer. *Gastric Cancer*. 2003;6(4):225–9.
- Cereceda Perez CN, Urbasos Pascual MI, Romero Castellanos C, Carreira Gomez C, Pinto Varela JM. Helical CT of the stomach: differentiation between benign and malignant pathologies, together with the staging of gastric carcinoma. *Rev Esp Enferm Dig*. 2002;94(10):601–12.

27. Chen BB, Liang PC, Liu KL, Hsiao JK, Huang JC, Wong JM, et al. Preoperative diagnosis of gastric tumors by three-dimensional multidetector row CT and double contrast barium meal study: correlation with surgical and histologic results. *J Formos Med Assoc.* 2007;106(11):943–52.
28. 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(2):472–82.
29. Chen J, Cheong JH, Yun MJ, Kim J, Lim JS, Hyung WJ, et al. Improvement in preoperative staging of gastric adenocarcinoma with positron emission tomography. *Cancer.* 2005;103(11):2383–90.
30. D'Elia F, Zingarelli A, Palli D, Grani M. Hydro-dynamic CT preoperative staging of gastric cancer: correlation with pathological findings. A prospective study of 107 cases. *Eur Radiol.* 2000;10(12):1877–85.
31. Dux M, Richter GM, Hansmann J, Kuntz C, Kauffmann GW. Helical hydro-CT for diagnosis and staging of gastric carcinoma. *J Comput Assist Tomogr.* 1999;23(6):913–22.
32. Gamon Giner R, Escrig Sos J, Salvador Sanchis JL, Ruiz del Castillo J, Garcia Vila JH, Marcote Valdivieso E. Helical CT evaluation in the preoperative staging of gastric adenocarcinoma. *Rev Esp Enferm Dig.* 2002;94(10):593–600.
33. Habermann CR, Weiss F, Riecken R, Honarparshah H, Bohnacker S, Staedtler C, et al. Preoperative staging of gastric adenocarcinoma: comparison of helical CT and endoscopic US. *Radiology.* 2004;230(2):465–71.
34. Hur J, Park MS, Lee JH, Lim JS, Yu JS, Hong YJ, et al. Diagnostic accuracy of multidetector row computed tomography in T- and N staging of gastric cancer with histopathologic correlation. *J Comput Assist Tomogr.* 2006;30(3):372–7.
35. Kadowaki K, Murakami T, Yoshioka H, Kim T, Takahashi S, Tomoda K, et al. Helical CT imaging of gastric cancer: normal wall appearance and the potential for staging. *Radiat Med.* 2000;18(1):47–54.
36. 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(1):14–24.
37. Kayaalp C, Arda K, Orug T, Ozcay N. Value of computed tomography in addition to ultrasound for preoperative staging of gastric cancer. *Eur J Surg Oncol.* 2002;28(5):540–3.
38. Kim HJ, Kim AY, Oh ST, Kim JS, Kim KW, Kim PN, et al. Gastric cancer staging at multi-detector row CT gastrotomography: comparison of transverse and volumetric CT scanning. *Radiology.* 2005;236(3):879–85.
39. Kumano S, Murakami T, Kim T, Hori M, Iannaccone R, Nakata S, et al. T staging of gastric cancer: role of multi-detector row CT. *Radiology.* 2005;237(3):961–6.
40. Lee DH, Ko YT. Advanced gastric carcinoma: the role of three-dimensional and axial imaging by spiral CT. *Abdom Imaging.* 1999;24(2):111–6.
41. Lee DH, Seo TS, Ko YT. Spiral CT of the gastric carcinoma: staging and enhancement pattern. *Clin Imaging.* 2001;25(1):32–7.
42. Lee DH, Ko YT, Park SJ, Lim JW. Comparison of hydro-US and spiral CT in the staging of gastric cancer. *Clin Imaging.* 2001;25(3):181–6.
43. Liao SR, Dai Y, Huo L, Yan K, Zhang L, Zhang H, et al. Transabdominal ultrasonography in preoperative staging of gastric cancer. *World J Gastroenterol.* 2004;10(23):3399–404.
44. Lim JS, Kim MJ, Yun MJ, Oh YT, Kim JH, Hwang HS, et al. Comparison of CT and 18F-FDG PET for detecting peritoneal metastasis on the preoperative evaluation for gastric carcinoma. *Korean J Radiol.* 2006;7(4):249–56.
45. Mochiki E, Kuwano H, Katoh H, Asao T, Oriuchi N, Endo K. Evaluation of 18F-2-deoxy-2-fluoro-D-glucose positron emission tomography for gastric cancer. *World J Surg.* 2004;28(3):247–53.
46. Polkowski M, Palucki J, Wronska E, Szawlowski A, Nasierowska-Guttmejer A, Butruk E. Endosonography versus helical computed tomography for locoregional staging of gastric cancer. *Endoscopy.* 2004;36(7):617–23.
47. Shimizu K, Ito K, Matsunaga N, Shimizu A, Kawakami Y. Diagnosis of gastric cancer with MDCT using the water-filling method and multiplanar reconstruction: CT-histologic correlation. *AJR Am J Roentgenol.* 2005;185(5):1152–8.
48. Shinohara T, Ohyama S, Yamaguchi T, Muto T, Kohno A, Ogura T, et al. Preoperative TNM staging of advanced gastric cancer with multi-detector row computed tomography. *Jpn Med Assoc J.* 2005;48(4):175–82.
49. Sohn KM, Lee JM, Lee SY, Ahn BY, Park SM, Kim KM. Comparing MR imaging and CT in the staging of gastric carcinoma. *AJR Am J Roentgenol.* 2000;174(6):1551–7.
50. Tian J, Chen L, Wei B, Shao M, Ding Y, Yin D, et al. The value of vesicant 18F-fluorodeoxyglucose positron emission tomography (18F-FDG PET) in gastric malignancies. *Nucl Med Commun.* 2004;25(8):825–31.
51. Wang CK, Kuo YT, Liu GC, Tsai KB, Huang YS. Dynamic contrast-enhanced subtraction and delayed MRI of gastric tumors: radiologic-pathologic correlation. *J Comput Assist Tomogr.* 2000;24(6):872–7.
52. Wei WZ, Yu JP, Li J, Liu CS, Zheng XH. Evaluation of contrast-enhanced helical hydro-CT in staging gastric cancer. *World J Gastroenterol.* 2005;11(29):4592–5.
53. 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(3):205–14.
54. Yang DM, Kim HC, Jin W, Ryu CW, Kang JH, Park CH, et al. 64 multidetector-row computed tomography for preoperative evaluation of gastric cancer: histological correlation. *J Comput Assist Tomogr.* 2007;31(1):98–103.
55. Yang QM, Kawamura T, Itoh H, Bando E, Nemoto M, Akamoto S, et al. Is PET-CT suitable for predicting lymph node status for gastric cancer? *Hepatogastroenterology.* 2008;55(82–83):782–5.
56. Yun M, Lim JS, Noh SH, Hyung WJ, Cheong JH, Bong JK, et al. Lymph node staging of gastric cancer using (18)F-FDG PET: a comparison study with CT. *J Nucl Med.* 2005;46(10):1582–8.
57. Hundt W, Braunschweig R, Reiser M. Assessment of gastric cancer: value of breathhold technique and two-phase spiral CT. *Eur Radiol.* 1999;9(1):68–72.
58. Kim YN, Choi D, Kim SH, Kim MJ, Lee SJ, Lee WJ, et al. Gastric cancer staging at isotropic MDCT including coronal and sagittal MPR images: endoscopically diagnosed early vs advanced gastric cancer. *Abdom Imaging.* 2009;34(1):26–34.
59. Takao M, Fukuda T, Iwanaga S, Hayashi K, Kusano H, Okudaira S. Gastric cancer: evaluation of triphasic spiral CT and radiologic-pathologic correlation. *J Comput Assist Tomogr.* 1998;22(2):288–94.
60. Kim SK, Kang KW, Lee JS, Kim HK, Chang HJ, Choi JY, et al. Assessment of lymph node metastases using 18F-FDG PET in patients with advanced gastric cancer. *Eur J Nucl Med Mol Imaging.* 2006;33(2):148–55.
61. Mukai K, Ishida Y, Okajima K, Isozaki H, Morimoto T, Nishiyama S. Usefulness of preoperative FDG-PET for detection of gastric cancer. *Gastric Cancer.* 2006;9(3):192–6.
62. Stahl A, Ott K, Weber WA, Becker K, Link T, Siewert JR, et al. FDG PET imaging of locally advanced gastric carcinomas: correlation with endoscopic and histopathological findings. *Eur J Nucl Med Mol Imaging.* 2003;30(2):288–95.

63. Blackshaw GR, Stephens MR, Lewis WG, Boyce J, Barry JD, Edwards P, et al. Progressive CT system technology and experience improve the perceived preoperative stage of gastric cancer. *Gastric Cancer*. 2005;8(1):29–34.
64. Bergman JJ, Fockens P. Endoscopic ultrasonography in patients with gastro-esophageal cancer. *Eur J Ultrasound*. 1999;10(2–3):127–38.
65. Cardoso R, Coburn N, Seevaratnam R, Sutradhar R, Law C, Lourenco L, et al. A systematic review and meta-analysis on the utility of EUS for pre-operative staging for early and advanced gastric cancer. *Gastric Cancer*. 2011; Accepted.
66. Kwee RM, Kwee TC. Imaging in assessing lymph node status in gastric cancer. *Gastric Cancer*. 2009;12(1):6–22.
67. Van Binnebeek S, Karges W, Mottaghy FM. Functional imaging of neuroendocrine tumors. *Methods Mol Biol*. 2011;727:105–22.
68. Ward J. New MR techniques for the detection of liver metastases. *Cancer Imaging*. 2006;6:33–42.
69. Leake P-A, Cardoso R, Seevaratnam R, Lourenco L, Helyer L, Mahar A, et al. A systematic review of the accuracy and indications for diagnostic laparoscopy prior to curative-intent resection of gastric cancer. *Gastric Cancer*. 2011. doi:[10.1007/s10120-011-0047-z](https://doi.org/10.1007/s10120-011-0047-z).
70. van Vliet EP, Steyerberg EW, Eijkemans MJ, Kuipers EJ, Siersema PD. Detection of distant metastases in patients with oesophageal or gastric cardia cancer: a diagnostic decision analysis. *Br J Cancer*. 2007;97(7):868–76.
71. Callaway MP, Bailey D. Staging computed tomography in upper GI malignancy. A survey of the 5 cancer networks covered by the South West Cancer Intelligence Service. *Clin Radiol*. 2005;60(7):794–800.