EUS for the staging of gastric cancer: a meta-analysis

Simone Mocellin, MD, PhD, Alberto Marchet, MD, Donato Nitti, MD
Padova, Italy

Background: The role of EUS in the locoregional staging of gastric carcinoma is undefined.

Objective: We aimed to comprehensively review and quantitatively summarize the available evidence on the staging performance of EUS.

Design: We systematically searched the MEDLINE, Cochrane, CANCERLIT, and EMBASE databases for relevant studies published until July 2010.

Setting: Formal meta-analysis of diagnostic accuracy parameters was performed by using a bivariate random-effects model.

Patients: Fifty-four studies enrolling 5601 patients with gastric cancer undergoing disease staging with EUS were eligible for the meta-analysis.

Main Outcome Measurements: EUS staging accuracy across eligible studies was measured by computing overall sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), and diagnostic odds ratio (DOR).

Results: EUS can differentiate T1-2 from T3-4 gastric cancer with high accuracy, with overall sensitivity, specificity, PLR, NLR, and DOR of 0.86 (95% CI, 0.81-0.90), 0.91 (95% CI, 0.89-0.93), 9.8 (95% CI, 7.5-12.8), 0.15 (95% CI, 0.11-0.21), and 65 (95% CI, 41-105), respectively. In contrast, the diagnostic performance of EUS for lymph node status is less reliable, with overall sensitivity, specificity, PLR, NLR, and DOR of 0.69 (95% CI, 0.63-0.74), 0.84 (95% CI, 0.81-0.88), 4.4 (95% CI, 3.6-5.4), 0.37 (95% CI, 0.32-0.44), and 12 (95% CI, 9-16), respectively. Results regarding single T categories (including T1 substages) and Bayesian nomograms to calculate posttest probabilities for any target condition prevalence are also provided.

Limitations: Statistical heterogeneity was generally high; unfortunately, subgroup analysis did not identify a consistent source of the heterogeneity.

Conclusions: Our results support the use of EUS for the locoregional staging of gastric cancer, which can affect the therapeutic management of these patients. However, clinicians must be aware of the performance limits of this staging tool. (Gastrointest Endosc 2011;73:1122-34.)

Despite its decreasing incidence in Western countries, gastric carcinoma remains the second leading cause of cancer deaths worldwide.1,2 In the United States, 21,000 new cases of this malignancy are estimated to occur in 2011, leading to 10,500 expected deaths.3 Radical surgery still represents the mainstay of treatment with curative intent.4,5 However, new approaches are gaining importance in the therapeutic management of these patients. For instance, EMR is proposed as an alternative to surgery for patients with early gastric cancer in...
the presence of favorable prognostic features (eg, histologically well-differentiated carcinoma limited to the mucosa, diameter <2 cm, absence of ulceration). Moreover, different adjuvant and neoadjuvant chemotherapy (with or without with radiotherapy) regimens have been shown to provide significant survival advantage to patients with advanced gastric cancer.

These strategies require reliable staging procedures to ensure the most appropriate (ie, with the highest therapeutic index, the ratio between efficacy and toxicity) treatment for each patient, according to the principles of personalized medicine.

In this regard, one of the most studied tools for the locoregional staging of gastric carcinoma is EUS. This endoscopy-based diagnostic device, which was introduced in the clinical setting in the 1980s, can both distinguish the different layers that compose the gastric wall and visualize the perigastric lymph nodes by means of a miniaturized US probe. Based on numerous reports published over more than 2 decades, EUS is often advocated as the best available method for the locoregional staging of gastric cancer; however, findings are heterogeneous (eg, sensitivity and specificity values range from 50% to 100%), and the comparison with more recent diagnostic techniques (eg, CT, magnetic resonance imaging) is difficult to perform because of the paucity of comparative studies conducted to date. In addition, although thousands of gastric cancer patients have been enrolled in EUS-based studies, no formal quantitative review of the available evidence has been published that comprehensively examines the staging performance of EUS by using the most appropriate statistical tools for the meta-analysis of diagnostic accuracy data.

In this work, we tried to fill this gap in the international literature by performing a formal systematic review and meta-analysis of the available evidence on this subject. Our aim was to provide readers with both a comprehensive overview and a quantitative analysis of the published data regarding the ability of EUS to identify primary tumor depth and regional lymph node status in patients with gastric carcinoma.

MATERIALS AND METHODS

Search strategy

We performed a comprehensive search of the English-language literature to identify articles that examined the diagnostic accuracy of EUS (the index test) in the evaluation of primary tumor depth of invasion (according to the American Joint Cancer Committee/Union Internationale Contre le Cancer T categories [T1, T2, T3, T4] as well as in terms of T1 subcategories mucosa [T1m] and submucosa [T1sm]) and regional lymph node status (metastatic [N+] vs disease free [N0]) by using histopathology as the reference standard.

We systematically searched the MEDLINE, Cochrane, CANCERLIT, and EMBASE databases for studies published until July 2010, by using the following search terms: “gastric” (or “stomach”), “cancer” (or “carcinoma”), “endoscopic” (or “endoscopy”), “ultrasound” (or “ultrasonography”), “endosonographic,” and “EUS.” We searched for additional references by cross-checking bibliographies of retrieved full-text papers.

Study selection and data extraction

We included studies that met all of the following inclusion criteria: (1) a minimal sample size of 10 patients with histologically proven primary carcinoma of the stomach; (2) evaluation of EUS compared with histopathology of primary tumor and lymph nodes; and (3) sufficient data to construct a 2 × 2 confusion matrix such that the cells in the table could be labeled as true positive, false positive, true negative, and false negative.

We excluded studies having possible overlap with the selected studies (ie, studies from the same study group, institution, and period of inclusion).

The index test was evaluated for the following diagnostic categories: (A) primary tumor depth: T1 (tumor limited to the mucosa and submucosa) versus non-T1 categories, T2 (tumor invading the muscularis propria ± subserosa) versus non-T2 categories, T3 (tumor invading the serosa) versus non-T3 categories, T4 (tumor infiltrating adjacent structures) versus non-T4 categories, T1-2 (tumor invading the gastric wall up to the muscularis propria/subserosa) versus T3-4 (tumor infiltrating the serosa or adjacent organs), T1m (T1 tumor limited to the mucosa layer) versus T1sm (T1 tumor involving the submucosa layer); (B) regional lymph node status: N+ (metastatic lymph nodes) versus N0 (disease-free lymph nodes). Because the TNM staging classification changed over time, it was not possible to assess the diagnostic accuracy of EUS across different N categories (N1, N2, N3).

The selection of studies for this meta-analysis was done by 2 independent reviewers (S.M., A.M.). Differences between the 2 reviewers were resolved by having a third reviewer (D.N.) assess the full article: the decision about whether to include the article was made by consensus. Relevant data were extracted from the articles selected for inclusion in the meta-analysis. Data quality was assessed by using a standard procedure according to the Quality...
Assessment of Diagnostic Accuracy Studies (QUADAS) criteria.

In addition to these data, we also recorded the following information for each study: overall study characteristics (including the first author, country, language, and date of publication) and characteristics regarding the index test (including type of echoendoscope [radial vs linear array transducer technology], US frequency, and EUS criteria for tumor depth and lymph node status).

In case of missing data, we requested information from the authors of the study.

When raw data were presented in 3 × 3 or 4 × 4 tables (eg, when the tumor depth [T] or lymph node stage was defined by >2 categories), we reconstructed 2 × 2 contingency tables by considering a given T (or any N-positive category) as the positive state to be distinguished from the other T categories (or from the N-negative cases).

Statistical analysis

As currently recommended for meta-analysis of diagnostic accuracy studies,27-29 we used a bivariate random-effects approach to obtain weighted overall estimates of the sensitivity (rate of cases correctly classified as diseased or true positive rate) and specificity (rate of cases correctly classified as healthy or true negative rate) of EUS in diagnosing both tumor depth infiltration and lymph node status.

This approach assumes a bivariate distribution for the logit-transformed values of paired sensitivity and specificity. Overall sensitivity and specificity and their 95% confidence intervals were calculated based on the binomial distributions of the true positives and true negatives. Besides accounting for study size and between-study heterogeneity, the bivariate model adjusts for the frequently observed negative correlation between the sensitivity and the specificity of the index test (threshold effect). An additional advantage of using the bivariate model is that the bivariate nature of the original data can be maintained throughout the analysis, allowing the generation of reliable summary estimates of sensitivity and specificity.

A summary receiver-operating characteristic (SROC) curve was constructed as a way to summarize the true-positive and false-positive rates from different studies. The SROC summary point represents the combination of the average sensitivity and specificity and is provided along with a 95% confidence interval as well as prediction interval.

The clinical (or patient-relevant) utility of EUS was evaluated by using likelihood ratios (which state how positive test results in participants with disease compared with the odds of positive test results in those without disease, on a natural logarithm scale against 1/√ESS [effect size], weighting by ESS: P < .10 for the slope coefficient indicates significant asymmetry of the funnel plot.31

All statistical analyses were conducted by using STATA 11.0 (StataCorp LP, College Station, Tex). All statistical tests were 2 sided, and the significance level was set at 5% (except for funnel plot asymmetry).

RESULTS

Eligible studies and quality assessment

As shown in Figure 1, the literature search identified 54 eligible studies published between 1988 and 201032-85: their main characteristics are reported in Table 1. Overall, 5601 patients were enrolled in 16 different countries, with a mean of 104 patients per study (range 14-388 patients).

The quality of the eligible studies, as assessed according to the QUADAS criteria, is reported in Figure 2. The percentage of high-quality studies (ie, those for which a yes response applied) varied between 13% and 100% for each of the 14 items. For most QUADAS items (8/14), all studies were classified as high quality, whereas for 3 of them (ie, uncertain results reporting, reference standard interpreted blind to index test result, and time between index test and reference standard description), the proportion of high-quality studies was less than 50%.
Primary tumor depth (T Stage)

Except for the 6 reports that did not explicitly describe the criteria, the definition of T stage for the index test was uniform across studies and relied on the fact that EUS visualizes the gastric wall as a 5-layer structure: accordingly, T1, T2, T3, and T4 tumors are defined based on the destruction of layers 1-3, 1-4, and 1-5 and infiltration of adjacent organs, respectively.

We first performed a meta-analysis of the eligible studies reporting data on T stages grouped into T1-2 versus T3-4 tumors to evaluate the ability of EUS to discriminate between early to intermediate (T1-2) and advanced (T3-4) gastric carcinoma. To this aim, 41 studies were available, for a total of 3510 patients.

The sensitivity and specificity of the single studies as well as their combined values are displayed in Figure 3. The SROC curve along with the summary point and the 95% confidence and prediction intervals are shown in Figure 4. Overall, the sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), and DOR were 0.86 (95% CI, 0.81-0.90), 0.91 (95% CI, 0.89-0.93), 9.8 (95% CI, 7.5-12.8), 0.15 (95% CI, 0.11-0.21), and 65 (95% CI, 41-105), respectively.

The Fagan plot (Fig. 5) shows that EUS can be clinically informative because it increases the previous probability of being classified as T1, T2, T3, and T4, respectively, from 35%, 20%, 35%, and 10% (average prevalence of corresponding T categories) to 91%, 63%, 75%, and 76% when positive; moreover, it lowers the above probabilities, respectively, to 9%, 9%, 8%, and 4% when negative.

Similarly, with regard to the ability of EUS to differentiate T1 mucosal (T1m) from submucosal (T1sm) tumors, a summary of the meta-analysis findings is provided in Table 2. Fagan plot analysis showed that EUS is less likely to be clinically useful to identify T1m tumors because it increases the previous probability of being classified as T1m from 75% (average T1m prevalence) to 92% when positive, and it lowers this probability to 39% when negative.

Of note, publication bias could not be excluded for these data (regression test P value = .06).
<table>
<thead>
<tr>
<th>Reference</th>
<th>Patients</th>
<th>Country</th>
<th>Reference test</th>
<th>EUS type</th>
<th>EUS MHz</th>
<th>Stomach site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murata et al, 1988&lt;sup&gt;32&lt;/sup&gt;</td>
<td>146</td>
<td>Japan</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-10</td>
<td>Any</td>
</tr>
<tr>
<td>Tio et al, 1989&lt;sup&gt;33&lt;/sup&gt;</td>
<td>80</td>
<td>the Netherlands</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Akahoshi et al, 1991&lt;sup&gt;34&lt;/sup&gt;</td>
<td>74</td>
<td>Japan</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>12</td>
<td>Any</td>
</tr>
<tr>
<td>Botet et al, 1991&lt;sup&gt;35&lt;/sup&gt;</td>
<td>50</td>
<td>United States</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Saito et al, 1991&lt;sup&gt;36&lt;/sup&gt;</td>
<td>110</td>
<td>Japan</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Caletti et al, 1993&lt;sup&gt;37&lt;/sup&gt;</td>
<td>35</td>
<td>Italy</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Dittler et al, 1993&lt;sup&gt;38&lt;/sup&gt;</td>
<td>254</td>
<td>Germany</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Grimm et al, 1993&lt;sup&gt;39&lt;/sup&gt;</td>
<td>147</td>
<td>Germany</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5</td>
<td>Any</td>
</tr>
<tr>
<td>Ziegler et al, 1993&lt;sup&gt;40&lt;/sup&gt;</td>
<td>108</td>
<td>Germany</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-20</td>
<td>Any</td>
</tr>
<tr>
<td>Shimizu et al, 1994&lt;sup&gt;41&lt;/sup&gt;</td>
<td>128</td>
<td>Japan</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Francois et al, 1996&lt;sup&gt;42&lt;/sup&gt;</td>
<td>29</td>
<td>France</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Cardia</td>
</tr>
<tr>
<td>Massari et al, 1996&lt;sup&gt;43&lt;/sup&gt;</td>
<td>65</td>
<td>Italy</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Perng et al, 1996&lt;sup&gt;44&lt;/sup&gt;</td>
<td>69</td>
<td>Taiwan</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Hamada, et al, 1997&lt;sup&gt;45&lt;/sup&gt;</td>
<td>149</td>
<td>Japan</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-20</td>
<td>Any</td>
</tr>
<tr>
<td>Yanai et al, 1997&lt;sup&gt;46&lt;/sup&gt;</td>
<td>104</td>
<td>Japan</td>
<td>Pathology of EMR or specimen</td>
<td>Radial, linear</td>
<td>20</td>
<td>Any</td>
</tr>
<tr>
<td>Akahoshi et al, 1998&lt;sup&gt;47&lt;/sup&gt;</td>
<td>75</td>
<td>Japan</td>
<td>Pathology of EMR or specimen</td>
<td>Radial</td>
<td>15</td>
<td>Any</td>
</tr>
<tr>
<td>Hunerbein et al, 1998&lt;sup&gt;48&lt;/sup&gt;</td>
<td>22</td>
<td>Germany</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>12.5</td>
<td>Any</td>
</tr>
<tr>
<td>Wang et al, 1998&lt;sup&gt;49&lt;/sup&gt;</td>
<td>119</td>
<td>Taiwan</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Nakamura et al, 1999&lt;sup&gt;50&lt;/sup&gt;</td>
<td>31</td>
<td>Japan</td>
<td>Pathology of surgical specimen</td>
<td>NR</td>
<td>NR</td>
<td>Any</td>
</tr>
<tr>
<td>Okamura et al, 1999&lt;sup&gt;51&lt;/sup&gt;</td>
<td>46</td>
<td>Japan</td>
<td>Pathology of EMR or specimen</td>
<td>Radial</td>
<td>20</td>
<td>Any</td>
</tr>
<tr>
<td>Yanai et al, 1999&lt;sup&gt;52&lt;/sup&gt;</td>
<td>49</td>
<td>Japan</td>
<td>Pathology of EMR or specimen</td>
<td>Radial, linear</td>
<td>20</td>
<td>Any</td>
</tr>
<tr>
<td>Mancone et al, 2000&lt;sup&gt;53&lt;/sup&gt;</td>
<td>79</td>
<td>Italy</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Tseng et al, 2000&lt;sup&gt;54&lt;/sup&gt;</td>
<td>74</td>
<td>Taiwan</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Willis et al, 2000&lt;sup&gt;55&lt;/sup&gt;</td>
<td>116</td>
<td>Germany</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Chen et al, 2002&lt;sup&gt;56&lt;/sup&gt;</td>
<td>57</td>
<td>Taiwan</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-20</td>
<td>Any</td>
</tr>
<tr>
<td>Hizawa et al, 2002&lt;sup&gt;57&lt;/sup&gt;</td>
<td>220</td>
<td>Japan</td>
<td>Pathology of EMR or specimen</td>
<td>Radial</td>
<td>12-20</td>
<td>Any</td>
</tr>
<tr>
<td>Xi et al, 2003&lt;sup&gt;58&lt;/sup&gt;</td>
<td>32</td>
<td>China</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-20</td>
<td>Any</td>
</tr>
<tr>
<td>Bhandari et al, 2004&lt;sup&gt;59&lt;/sup&gt;</td>
<td>48</td>
<td>Korea</td>
<td>Pathology of EMR or specimen</td>
<td>Radial</td>
<td>20</td>
<td>Any</td>
</tr>
<tr>
<td>Habermann et al, 2004&lt;sup&gt;60&lt;/sup&gt;</td>
<td>51</td>
<td>Germany</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Hunerbein et al, 2004&lt;sup&gt;61&lt;/sup&gt;</td>
<td>49</td>
<td>Germany</td>
<td>Pathology of EMR or specimen</td>
<td>NR</td>
<td>12.5</td>
<td>Any</td>
</tr>
<tr>
<td>Javaid et al, 2004&lt;sup&gt;62&lt;/sup&gt;</td>
<td>112</td>
<td>India</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5</td>
<td>Any</td>
</tr>
<tr>
<td>Polkowski et al, 2004&lt;sup&gt;63&lt;/sup&gt;</td>
<td>88</td>
<td>Poland</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Shimoyama et al, 2004&lt;sup&gt;64&lt;/sup&gt;</td>
<td>45</td>
<td>Japan</td>
<td>Pathology of surgical specimen</td>
<td>Linear</td>
<td>7.5</td>
<td>Cardia</td>
</tr>
<tr>
<td>Pedrazzani et al, 2005&lt;sup&gt;65&lt;/sup&gt;</td>
<td>51</td>
<td>Italy</td>
<td>Pathology of surgical specimen</td>
<td>Linear</td>
<td>7.5</td>
<td>Cardia</td>
</tr>
<tr>
<td>Yoshida et al, 2005&lt;sup&gt;66&lt;/sup&gt;</td>
<td>293</td>
<td>Japan</td>
<td>Pathology of EMR or specimen</td>
<td>Radial</td>
<td>7.5-20</td>
<td>Any</td>
</tr>
<tr>
<td>Akashi et al, 2006&lt;sup&gt;67&lt;/sup&gt;</td>
<td>267</td>
<td>Japan</td>
<td>Pathology of EMR or specimen</td>
<td>NR</td>
<td>12-20</td>
<td>Any</td>
</tr>
<tr>
<td>Ang et al, 2006&lt;sup&gt;68&lt;/sup&gt;</td>
<td>57</td>
<td>Singapore</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
</tbody>
</table>
Figure 2. The quality of the eligible studies as assessed according to the 14 items included in the Quality Assessment of Diagnostic Accuracy Studies criteria. Dark blue bars indicate yes, light blue bars no, and green bars uncertain/unclear.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Patients</th>
<th>Country</th>
<th>Reference test</th>
<th>EUS type</th>
<th>EUS MHz</th>
<th>Stomach site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arocena et al, 2006</td>
<td>17</td>
<td>Spain</td>
<td>Pathology of surgical specimen</td>
<td>Linear</td>
<td>12.5</td>
<td>Any</td>
</tr>
<tr>
<td>Ganpathi et al, 2006</td>
<td>102</td>
<td>Singapore</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Potrc et al, 2006</td>
<td>82</td>
<td>Slovenia</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Tsendsuren et al, 2006</td>
<td>41</td>
<td>China</td>
<td>Pathology of surgical specimen</td>
<td>Linear</td>
<td>5-7.5</td>
<td>Any</td>
</tr>
<tr>
<td>Barbour et al, 2007</td>
<td>206</td>
<td>USA</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Cardia</td>
</tr>
<tr>
<td>Bentrem et al, 2007</td>
<td>218</td>
<td>United States</td>
<td>Pathology of surgical specimen</td>
<td>NR</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Tan et al, 2007</td>
<td>63</td>
<td>China</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-20</td>
<td>Any</td>
</tr>
<tr>
<td>Blackshaw et al, 2008</td>
<td>44</td>
<td>UK</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Cardia</td>
</tr>
<tr>
<td>Lok et al, 2008</td>
<td>75</td>
<td>Hong Kong</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>12-20</td>
<td>Any</td>
</tr>
<tr>
<td>Park et al, 2008</td>
<td>40</td>
<td>Korea</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Ahn et al, 2009</td>
<td>71</td>
<td>Korea</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>5-12</td>
<td>Any</td>
</tr>
<tr>
<td>Heye et al, 2009</td>
<td>14</td>
<td>Germany</td>
<td>Pathology of surgical specimen</td>
<td>NR</td>
<td>NR</td>
<td>Any</td>
</tr>
<tr>
<td>Mouri et al, 2009</td>
<td>222</td>
<td>Japan</td>
<td>Pathology of EMR or specimen</td>
<td>Radial</td>
<td>12-20</td>
<td>Any</td>
</tr>
<tr>
<td>Choi et al, 2010</td>
<td>388</td>
<td>Korea</td>
<td>Pathology of EMR or specimen</td>
<td>Radial</td>
<td>12</td>
<td>Any</td>
</tr>
<tr>
<td>Hwang et al, 2010</td>
<td>277</td>
<td>Korea</td>
<td>Pathology of EMR or specimen</td>
<td>Radial</td>
<td>5-20</td>
<td>Any</td>
</tr>
<tr>
<td>Kim et al, 2010</td>
<td>176</td>
<td>Korea</td>
<td>Pathology of EMR or specimen</td>
<td>Radial</td>
<td>20</td>
<td>Any</td>
</tr>
<tr>
<td>Repiso et al, 2010</td>
<td>36</td>
<td>Spain</td>
<td>Pathology of surgical specimen</td>
<td>Radial</td>
<td>7.5-20</td>
<td>Any</td>
</tr>
</tbody>
</table>

NR, Not reported.
Lymph node status (N Stage)

We then performed a meta-analysis of the eligible studies reporting data on N stage (positive vs negative) to evaluate the ability of EUS to diagnose regional lymph node status of patients with gastric carcinoma. To this aim, 39 studies were available, for a total of 3315 patients. Except for the 6 reports that did not describe the criteria, the EUS definition of N stage varied across studies; in particular, some investigators relied exclusively on lymph node size (≥8 mm), whereas most relied on lymph node morphology (e.g., sharp margin and hypoechoic pattern) coupled or not with size value.

Sensitivities and specificities of the single studies as well as their combined values are displayed in Figure 6. The SROC curve along with the summary point and its 95% confidence and prediction intervals are shown in Figure 7. Overall sensitivity, specificity, PLR, NLR, and DOR were 0.69 (95% CI, 0.63-0.74), 0.84 (95% CI, 0.81-0.88), 4.4 (95% CI, 3.6-5.4), 0.37 (95% CI, 0.32-0.44), and 12 (95% CI, 9-16), respectively.

The Fagan plot (Fig. 8) shows that EUS could be clinically informative because it increases the probability of being classified as N+ from 55% (average prevalence of N+ cases) to 84% when positive, and it lowers the same probability to 31% when negative.

Between-study heterogeneity was significant both for sensitivity ($I^2$: 81.5%, Q test $P < .0001$) and specificity ($I^2$: 72.4%, Q test $P < .0001$). The proportion of heterogeneity likely caused by the threshold effect was substantial (37%). Subgroup analysis showed higher sensitivity (0.72 [95% CI, 0.66-0.78] vs 0.58 [95% CI, 0.47-0.70]) and lower specificity (0.83 [95% CI, 0.78-0.87] vs 0.88 [95% CI, 0.83-0.93]) in studies with higher disease prevalence (lymph node positivity ≥50%); in contrast, studies using high-frequency EUS reported on average lower sensitivity (0.60 [95% CI, 0.49-0.71] vs 0.72 [95% CI, 0.67-0.78]) and higher specificity (0.88 [95% CI, 0.83-0.94] vs 0.83 [95% CI, 0.78-0.87]).

Regression testing for funnel plot asymmetry showed no evidence of statistically significant publication bias ($P = .93$).

**DISCUSSION**

This meta-analysis, the first to include such a large cohort of patients (N = 5601) and conducted with modern statistical methods, quantitatively summarizes the avail-

---

**Figure 3.** EUS diagnostic performance to distinguish T1-2 from T3-4 tumors: forest plot of the studies included in the meta-analysis.
able evidence of the diagnostic performance of EUS for the locoregional staging of gastric cancer.

We found that EUS evaluation is associated with clinically relevant high performance rates (Figs. 3-5) for differentiating between early to intermediate (T1-2) and advanced (T3-4) primary gastric tumors. In particular, the average positive and negative likelihood ratios of EUS (9.8 and 0.15, respectively) approach the desirable values (≥10 and ≤0.1, respectively) for optimal diagnostic tests. These figures, along with the Bayesian nomogram provided in Figure 5, support the use of EUS in the clinical setting, with special regard to the therapeutic management of patients who are to be selected for neoadjuvant treatments before surgery, as mentioned previously.

The diagnostic accuracy for the single T categories (T1, T2, T3, and T4) is overall less satisfactory (Table 2), although one could pinpoint that the identification of each single category is less important than the ability to differentiate T1-2 from T3-4 tumors from a practical viewpoint (ie, the therapeutic management of patients). In the T1 category, a clinically relevant issue is the differentiation of T1m from T1sm lesions, which would enable clinicians to better select patients who might benefit from EMR. Unfortunately, the average EUS performance in this subset of patients does not appear to be sufficiently informative on practical grounds, although the data on this subject are from a smaller number of studies, and publication bias could not be excluded in our meta-analysis (Table 2).

With regard to regional lymph node status, the ability of EUS to distinguish between positive and negative cases is unsatisfactory. However, although all diagnostic performance parameters are worse than those described for the T1-2/T3-4 staging issue, the Fagan plot (Fig. 8) shows that EUS could still be clinically informative because it increases the previous probability of being classified as N+ from 55% (average prevalence of N+ cases) to 84% when positive, and it lowers the same probability to 31% when negative. Accordingly, EUS can provide clinicians with an estimate of the risk of patients with lymph node metastatic disease; although this information is not as reliable as that supplied for primary tumor depth, it can still represent a further clue in selecting patients for preoperative treatments.

Overall, a common feature of our findings is that statistical heterogeneity was generally high, a very frequent observation in meta-analysis of diagnostic studies. Unfor-
TABLE 2. Meta-analysis of EUS diagnostic performance for T1, T2, T3, and T4 as well as T1 mucosal (T1m) gastric carcinoma

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T1m</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of studies</td>
<td>41</td>
<td>39</td>
<td>39</td>
<td>34</td>
<td>15</td>
</tr>
<tr>
<td>Patients</td>
<td>3866</td>
<td>3475</td>
<td>3444</td>
<td>3017</td>
<td>2318</td>
</tr>
<tr>
<td>Sensitivity (95% CI)</td>
<td>0.83 (0.77-0.88)</td>
<td>0.65 (0.57-0.72)</td>
<td>0.86 (0.83-0.89)</td>
<td>0.66 (0.52-0.77)</td>
<td>0.83 (0.76-0.89)</td>
</tr>
<tr>
<td>Specificity (95% CI)</td>
<td>0.96 (0.93-0.97)</td>
<td>0.91 (0.88-0.92)</td>
<td>0.85 (0.80-0.89)</td>
<td>0.98 (0.97-0.98)</td>
<td>0.79 (0.65-0.88)</td>
</tr>
<tr>
<td>PLR (95% CI)</td>
<td>19.8 (12.7-31.1)</td>
<td>6.9 (5.4-8.9)</td>
<td>5.7 (4.3-7.5)</td>
<td>28.1 (18.5-42.5)</td>
<td>3.9 (2.4-6.3)</td>
</tr>
<tr>
<td>NLR (95% CI)</td>
<td>0.18 (0.13-0.24)</td>
<td>0.39 (0.31-0.48)</td>
<td>0.16 (0.13-0.2)</td>
<td>0.35 (0.24-0.51)</td>
<td>0.21 (0.16-0.28)</td>
</tr>
<tr>
<td>DOR (95% CI)</td>
<td>112 (70-179)</td>
<td>18 (12-27)</td>
<td>35 (24-52)</td>
<td>80 (41-153)</td>
<td>19 (13-27)</td>
</tr>
</tbody>
</table>

Heterogeneity

<table>
<thead>
<tr>
<th></th>
<th>P, %</th>
<th>99</th>
<th>95</th>
<th>97</th>
<th>93</th>
<th>99</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q test P value</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Threshold effect, %

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>44</th>
<th>1</th>
<th>1</th>
<th>5</th>
<th>96</th>
</tr>
</thead>
</table>

Publication bias test

|        | P value | .80   | .16   | .92   | .30   | .06   |

CI, Confidence interval; PLR, positive likelihood ratio; NLR, negative likelihood ratio; DOR, diagnostic odds ratio.

Figure 6. EUS diagnostic performance to distinguish lymph node–positive from lymph node–negative tumors: forest plot of the studies included in the meta-analysis.
Fortunately, subgroup analysis did not identify a consistent source of heterogeneity, which appears to be linked at least in part to year of publication, a combination of threshold effects, disease prevalence and EUS frequency, and publication bias, depending on the type of diagnostic issue addressed (T1-2 vs T3-4, N+ vs N−, and T1m vs T1sm, respectively). Another potential source of heterogeneity is operator experience, an aspect often emphasized for US-based diagnostic imaging; however, this issue could not be computationally addressed in our meta-analysis because of the scarcity (or absence) of data reported by single studies on this subject. Based on these findings, no specific recommendation can be made on how to improve the standardization (and thus the consistency) of the EUS results.

With regard to the already existing literature on this subject, 2 systematic reviews (without meta-analysis) have been published on EUS and gastric cancer locoregional staging between 2007 and 2009. In the former, 22 which was dedicated to primary tumor depth, the authors conclude that EUS remains the first-choice imaging modality in T staging compared with magnetic resonance imaging and CT; in the latter, 18 which addressed the issue of lymph node status, the same investigators conclude that EUS (like magnetic resonance imaging and CT) cannot reliably be used to confirm or exclude lymph node metastasis in gastric cancer. These conclusions are similar to those that we have drawn, although our meta-analysis provides formal evidence to sustain these hypotheses and more importantly provides readers with a quantification of the average performance of this endoscopic tool. The latter aspect, along with the Bayesian nomograms, allows clinicians to get a precise sense of the risk of making errors (both in terms of false-positive and false-negative predictions) while using EUS, which ultimately can help them optimize the therapeutic management of patients based on probabilities and not generic dichotomic (works/does not work) expert opinions. In this respect, we strongly believe that the much-cited “personalized medicine” passes also through the use of the available tests based on the awareness of their predictive values and not just on the knowledge of their “raw” binary response (positive/negative).

Finally, between 2001 and 2008, 2 meta-analyses were published on both T staging and N staging of gastric cancer...
cancer with EUS. In the first, the authors conclude that EUS is more accurate for the staging of advanced rather than early disease; in the second, which also included esophageal carcinomas, the investigators state that EUS is highly effective in differentiating T1-2 from T3-4 tumors. However, these findings are not comparable to ours because of the fact that the methods used in that work (based on the Moses-Littenberg model) are no longer considered scientifically sound for the meta-analysis of diagnostic accuracy studies and because the number of included studies was much lower than what we could retrieve and analyze.

Taken together, our results support the use of EUS for the locoregional staging of gastric cancer, provided that clinicians are aware of the diagnostic performance limits of this tool, as outlined earlier. Technological improvements (in particular in the definition of lymph node status) such as those recently proposed may lead to better EUS performance rates and thus to the optimization of gastric cancer staging, which should ultimately ameliorate the clinical management of these patients.

ACKNOWLEDGMENTS

The authors thank Salvatore Correra for his help in retrieving the relevant articles. They also thank Dr. Marta Briarava for setting up the dedicated database used to perform this meta-analysis.

REFERENCES


