ORIGINAL ARTICLE

ueg journal WILEY

Expert assessment of infiltration depth and recommendation of endoscopic resection technique in early Barrett cancer

Fadi Younis[1,2] | Thomas Rösch[1] | Torsten Beyna[3] | Alanna Ebigbo[4] | Siegbert Faiss[5] | Andrea May[6] | Oliver Pech[7] | Philip Dautel[1] | Mario Anders[8] | Till Clauditz[9] | Katharina Zimmermann-Fraedrich[1] | Susanne Sehner[10] | Guido Schachschal[1]

[1]Department of Interdisciplinary Endoscopy, University Hospital Hamburg-Eppendorf, Hamburg, Germany

[2]Department of Gastroenterology and Hepatology, Tel Aviv Sourasky Medical Center, and Tel-Aviv University, Tel Aviv, Israel

[3]Department of Internal Medicine and Gastroenterology, Evangelisches Krankenhaus, Düsseldorf, Germany

[4]Department of Gastroenterology, Klinikum Augsburg, Augsburg, Germany

[5]Department of Gastroenterology, Sana Klinikum Lichtenberg, Berlin, Germany

[6]Department of Gastroenterology, Hepatology, Oncology and Pneumology, Asklepios Paulinen Hospital, Wiesbaden, Germany

[7]Department of Gastroenterology and Hepatology, Krankenhaus Barmherzige Brüder, Regensburg, Germany

[8]Department of Gastroenterology and Interdisciplinary Endoscopy, Vivantes Auguste Viktoria Hospital, Berlin, Germany

[9]Institute of Pathology, University Hospital Hamburg-Eppendorf, Hamburg, Germany

[10]Department of Medical Biometry and Epidemiology, University Hospital Hamburg-Eppendorf, Hamburg, Germany

Correspondence

Thomas Rösch, Department of Interdisciplinary Endoscopy, University Hospital Hamburg Eppendorf, Martinistr. 52, Hamburg 20246, Germany.
Email: t.roesch@uke.de

Abstract

Background: Early Barrett cancer can be curatively treated by endoscopic resection. The choice of the resection technique, however—endoscopic mucosal resection (EMR) or submucosal dissection (ESD)—largely depends on the assumed infiltration depth as judged by the endoscopist. However, the accuracy of endoscopic diagnosis of the degree of cancer infiltration is not known.

Methods: Three to four high-quality images (both in overview and close-up) from 202 of early Barrett esophagus cancer cases (82% men, mean age 66.9 years) were selected from our endoscopy database (73.3% stage T1a and 26.7% in stage T1b). Images were shown to 9 Barrett esophagus experts, with patients' clinical data (age, sex, Barrett esophagus length) and biopsy results. The experts were asked to predict infiltration depth (T1b vs. T1a), and to suggest the appropriate endoscopic resection technique (EMR or ESD, or surgery). Interobserver variability (kappa values) was also determined for these parameters.

Results: Overall positive (PPV) and negative predictive values (NPV) to diagnose T1b versus T1a infiltration were 40.7% (95% CI: 36.7, 44.8) and 79.8% (95% CI: 77.5, 81.9), respectively; kappa value was 0.41. Paris classification (kappa 0.51) and suggested treatment also varied between experts. In a post hoc analysis, only the correlation between lesions classified as invisible or flat according to the Paris classification (IIB; 25% of all cases) and the suggested resection technique was better: In this subgroup, EMR was recommended in >80% of cases, with a high complete (basal R0) resection rate (mean of 88.1%).

Conclusions: Precise endoscopic distinction between mucosal and submucosal involvement of Barrett esophagus cancer by experts as a basis for choosing the resection technique has limited predictive values and high interobserver variability. It seems that mainly invisible/flat lesions may result in good resection outcomes

when treated by EMR, but this stratification strategy has to be assessed in further studies.

KEYWORDS
Barrett esophagus, early cancer, endoscopic resection, interobserver variability, staging

INTRODUCTION

Early neoplasia arising from Barrett's esophagus (BE) can be treated curatively by endoscopic resection in combination with subsequent ablation for the remaining non-neoplastic BE.[1,2] Complete resection is primarily defined as histologically complete (R0) resection, esp. with regards to the basal resection margin, but "curative" resection depends on oncologic risk criteria, which may require further management such as subsequent surgery in case of high risk lesions even if the lesion had been completely (R0) resected by endoscopy.[3] These risk categories are distinguished by certain histologic criteria such as grading, infiltration depth, and presence of lymphovascular involvement.[1,4,5]

This discrepancy between complete and curative resection can be found throughout the literature for early cancers in the GI tract. It may have decisive influence on the choice of the resection technique, since the advantage of a higher rate of complete resection may not translate into a similar difference in curative resection rates. Therefore, it is still unclear which of the two competing resection methods available, namely endoscopic mucosal resection (EMR) and endoscopic submucosal dissection (ESD), should be used in BE neoplasia. There is no evidence on the comparative oncologic outcome from randomized comparative studies, so that daily practice may be guided by guidelines: These suggest EMR for less and ESD for more advanced lesions, but this distinction is mostly based on histology of the resection specimen using differentiated criteria which are not evident from prior biopsy but can only be elucidated from the resection specimen post hoc.[2,3,6,7] However, in daily practice, the endoscopist has to decide upfront from the endoscopic aspect and to judge how deep this lesion may infiltrate to choose the appropriate resection technique.

Since the pre-procedural stratification of patients to either EMR or ESD mostly depends on the macroscopic aspect, ideally, there should be a good correlation of endoscopic morphology and final histology: In that case, EMR would be used for flat lesions limited to the mucosa (T1a), while ESD is used for early tumors with submucosal involvement (T1b). Other imaging methods such as endoscopic ultrasound[8,9] and computed tomography[10] were shown not to be particularly useful either. Data on the accuracy of endoscopists to correctly classify lesions by endoscopy with regard to their infiltration depth are very limited.[11] We therefore conducted an interobserver study on images of a large number of patients with T1 BE carcinoma, including therapeutic recommendations of ESD versus EMR.

Key summary

Established knowledge on this subject

- Endoscopic resection of early Barrett cancer has been established as the standard for certain histologic (low risk) subtypes. The choice of the resection method is less clear and not evidence-based, but guidelines recommend piecemeal endoscopic mucosal resection (EMR) for pre-neoplastic and early mucosal lesions and en bloc endoscopic submucosal dissection (ESD) for more advanced early tumors. Such an approach however rests on the endoscopist's ability to differentiate between these two stages of early neoplasia

What are the significant and/or new findings?

- From our endoscopic database, we selected representative images of 202 T1a and T1b early Barrett cancer treated endoscopically and showed them to 9 experienced endoscopists in the area of Barrett neoplasia. These experts failed in reliably differentiating between T1a and T1b with a low kappa value (0.41) and also differed in their recommendation of the endoscopic resection method. In the subgroup of invisible or flat neoplasia, results were somewhat better, with better resection outcomes when treated by EMR. Further studies have to analyze whether such an allocation strategy will work, that is, to limit EMR to these lesions and treat the remaining ones by ESD

PATIENTS AND METHODS

Patients and data retrieval

Records from all patients treated endoscopically (ESD or EMR) with BE associated neoplasia between 2009 and 2022 were retrieved from the endoscopy database at our center (Department of Interdisciplinary Endoscopy at University Hospital Hamburg-Eppendorf). Only patients with post resection cancer histology and adequate lesion images were analyzed, see Figure S1. Expert gastrointestinal pathologists assessed the samples and reported the final diagnoses. The invasion depth of cancer was classified according to the Japanese

classification of esophageal cancer and was measured in microns,[12] differentiating between T1a cancer involving the mucosal layer (=T1m) only and T1b cancer with submucosal infiltration (=T1sm1<500 μm/sm2<1000 μm/sm3).

Image selection

Adequate images of the lesions were selected by a fellow with 1 year of intensive BE training who was not part of the panel of reviewers. All images were captured using a high-definition gastroscope (Olympus Corp. Europe, Hamburg, Germany), including assessments of dye- (acetic acid) or virtual based chromoendoscopy (NBI/BLI). From each resection case selected, the three to four best endoscopic images demonstrating the lesion in overview (including surrounding BE) and close up views were taken, following in-house standards to take images. 202 resection cases with 700 adequate lesion images (110 ESD with 361 images + 92 EMR with 339 images) were pooled in the high-quality format and inserted into a Microsoft PowerPoint slide presentation with a white background. Each case was numbered and randomized. Assigned file size and format were preserved during this process without compromising image quality.

Experts review and interpretation of images

These image cases as well as clinical (age, sex, BE length) and preprocedural biopsy information including cancer grading were reviewed by 9 BE experts (6 external and 3 internal who had done the resections, in one hospital), with experience in managing Barrett esophagus of at least 5 years and in >100 BE resections (50 ESD, 50 EMR) and had a track record of publications in the area of BE (see literature list in the Appendix). All experts were blinded to the final resection technique chosen and the final histopathological results. All reviewers received a standardized scoring sheet for each case (total 202) for the assessment of: Paris classification (including a schematic drawing for each case to remind the rater of the different morphologies), histopathological diagnosis T1a versus T1b, T1bsm1 or deeper, confidence level, recommended resection technique and reason for choosing of the technique (see **CRF** and Figures S1 and S2 in the Supporting Information S1). Criteria to diagnose T1b versus T1a were based on subjective assessment (tumor bulkiness or depression/ulceration indicating T1b, size >2 cm, lesion hardness etc.) since there are no established endoscopic imaging criteria published up to now.

Outcomes

The *main outcome* of the present study was positive (PPV) and negative predictive value (NPV) of experts to correctly diagnose stage T1b (submucosal) versus T1a (mucosal infiltration).

Secondary outcomes were:

- Subgrouping of T1b cases into T1b-sm1 (500 μm or less) versus deeper infiltration (T1b sm2/3), that is, for expert assessment the differentiation between T1b sm2/3 versus T1b sm1 plus T1a. This reflects current guidelines.[3]
- The choice of the resection technique, EMR or ESD, or, if deemed necessary, surgery, as recommended by experts on the basis of the above images.
- Paris classification by experts.[13,14]
- Interobserver variability (kappa values) of the above parameters using the Fleiss-Kappa method. According to Landis & Koch[15] we adhered to the following categorization: 0.20: slight; 0.21–0.40: fair; 0.41–0.60: moderate; 0.61–0.90: substantial; 0.81–1: almost perfect.

Statistical analysis and IRB

Baseline characteristics are reported as mean with standard deviation (SD) for continuous variables, median with 1st and 3rd quartile (IQR) for categorical variables, and frequencies and percentages for binary variables. Accuracy, predictive values, sensitivity and specificity were calculated stratified by experts and reported with 95% confidence intervals (95% CI). Overall sensitivity and specificity were estimated using a random-effects meta-analysis for diagnostic test accuracy and visualized within forest plots. In addition, interobserver agreement was estimated by calculating Fleiss' kappa-statistic with bias-corrected bootstrapped 95%-confidence intervals. Analyses were performed using Stata 18[16] (StataCorp. 2023. Stata Statistical Software: Release 18. College Station, TX: StataCorp LLC).

The clinical study was approved by the local ethics committee for the publication of quality assurance data (Hamburg Chamber of Physicians, PV 2022-300223-WF). Images were pseudonymized for external assessment.

RESULTS

Patient characteristics

The flow sheet of the selected study cases is shown in Figure S1. Of 560 cases extracted from the database, 149 were excluded since they were part of a current randomized trial of ESD versus EMR (NCT03427346), 116 had final histology other than cancer in the resection specimen (including HGD only), and the remaining excluded cases had other reasons such as mainly insufficient photographic quality. Only cases with endoscopic resection were included, but 44/202 (21.8%) cases were operated secondarily, mostly for oncologic reasons (R1 resection and/or high-risk lesions, see below). Characteristics of the 202 patients are shown in Table 1 and Table S1. Age (67 years) and sex distribution (>80% male patients) as well as BE length (55% long BE) were within expected ranges. Resection techniques chosen in clinical practice at

TABLE 1 Baseline demographics of included patients, resection technique and results (for further details see Table S1).

Patient and Barrett details	All cases (202)	T1a cases (148)	T1b cases (54)[a]
Sex, male, n (%)	166 (82.2%)	122 (82.4%)	44 (81.5%)
Age, years, mean (±SD)/median (IQR)	66.9 + 10.7	66.8 (10.6)	67.2 (11.2)
	68 (59–75)	68 (59–75)	70 (59–76)
Number of images, median (IQR)	4 (3–4)	4 (3–4)	4 (3–4)
Circular extent (C), cm, median (IQR)	1.0 (0.0–5.0)	1.0 (0.0–4.0)	2.0 (0.0–6.0)
Tongue extent (M), cm, median (IQR)	4.0 (2.0–7.0)	3.0 (2.0–7.0)	5.0 (2.0–7.0)
LSB (M = 3 cm or more), n (%)	109 (54.8%)	76 (52.4%)	33 (61.1%)
Actual resection technique			
EMR	92	73	19
ESD	110	75	35
Complete/R0 resection			
Deep margins	168 (83.2%)	135 (91.2%)	33 (61.1%)
Deep and lateral margins	164 (81.2%)	131 (88.5%)[b]	33 (61.1%)

[a]of those 35 were T1sm1 (500 m or less) and 19 T1sm2/3.
[b]4 patients with T1a had positive lateral margins.
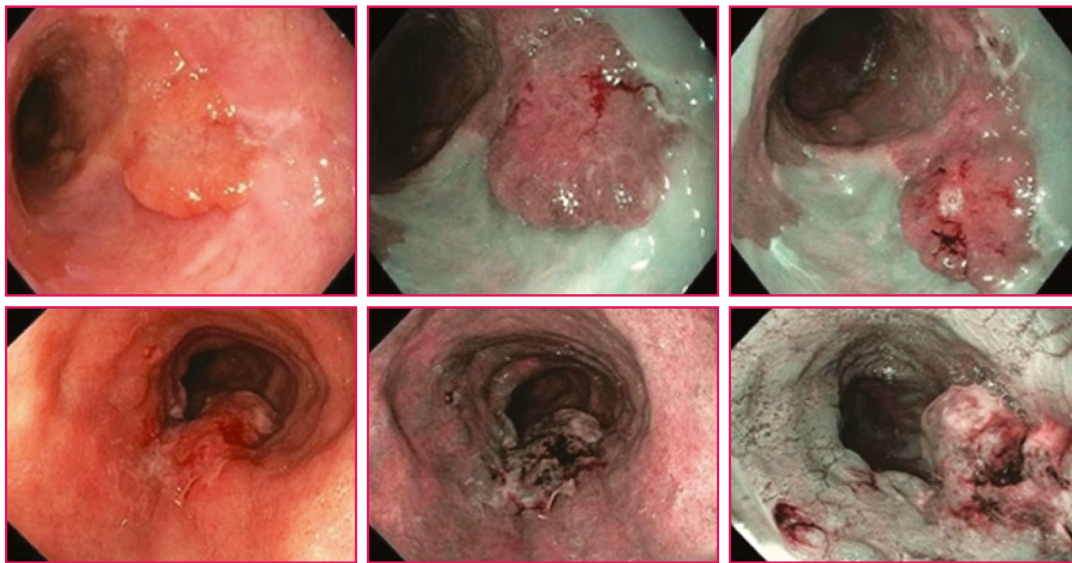

FIGURE 1 Examples of two different lesions resected with endoscopic submucosal dissection. Upper row shows three different lesion images of mucosal cancer (T1a), lower row shows three different lesion images of submucosal invasive cancer (T1b).

our center were almost equally distributed. Table S2 shows the overall results for all parameters summarized as classified by all 9 raters. Figure 1 shows endoscopic image examples for histologic stage T1a and stage T1b.

Lesion assessment for infiltration depth (T1b vs. T1a; main outcome)

Figure 2 (and Table S3) show the results for the staging classification by the 9 experts with regard to histologic tumor infiltration depth, which they were asked to predict. Accuracy and positive and negative predictive values to recognize T1b versus T1a as main outcome parameters were variable: Accuracy rates were between 60.4 (95% CI: 53.3, 67.2) and 73.1% (95% CI: 66.3, 79.2). Sensitivity and specificity are also shown in Figure 2. Kappa value was 0.41 overall (95% CI: 0.34, 0.47) and thus on the border between low and moderate.

Furthermore, we (as a *secondary outcome*) aimed at subdifferentiating the T1b cases into sm1 versus sm2/3, thus shifting T1sm1 histology into the low-risk group as per recent guidelines[3]; results are also shown in Figure 2 and Table S3. For the new cut-off (T1b sm2/3

T1b vs T1a Kappa over all experts 0.41 (95% CI: 0.35, 0.47)		
Expert	Positive predictive value (95% CI)	Negative predictive value (95% CI)
1	48.9% (33.7%, 64.2%)	80.3% (73.0%, 86.3%)
2	40.8% (27.0%, 55.8%)	79.1% (71.4%, 85.6%)
3	35.2% (25.3%, 46.1%)	79.8% (71.3%, 86.8%)
4	37.7% (26.9%, 49.4%)	80.6% (72.6%, 87.2%)
5	43.1% (30.8%, 56.0%)	80.9% (73.3%, 87.1%)
6	48.6% (31.4%, 66.0%)	77.8% (70.8%, 83.9%)
7	38.9% (25.9%, 53.1%)	77.7% (70.1%, 84.1%)
8	43.7% (33.9%, 56.0%)	82.3% (74.6%, 88.2%)
9	37.8% (26.8%, 49.9%)	80.3% (72.3%, 86.8%)
Overall	40.7% (36.7%, 44.8%)	79.8% (77.5%, 81.9%)

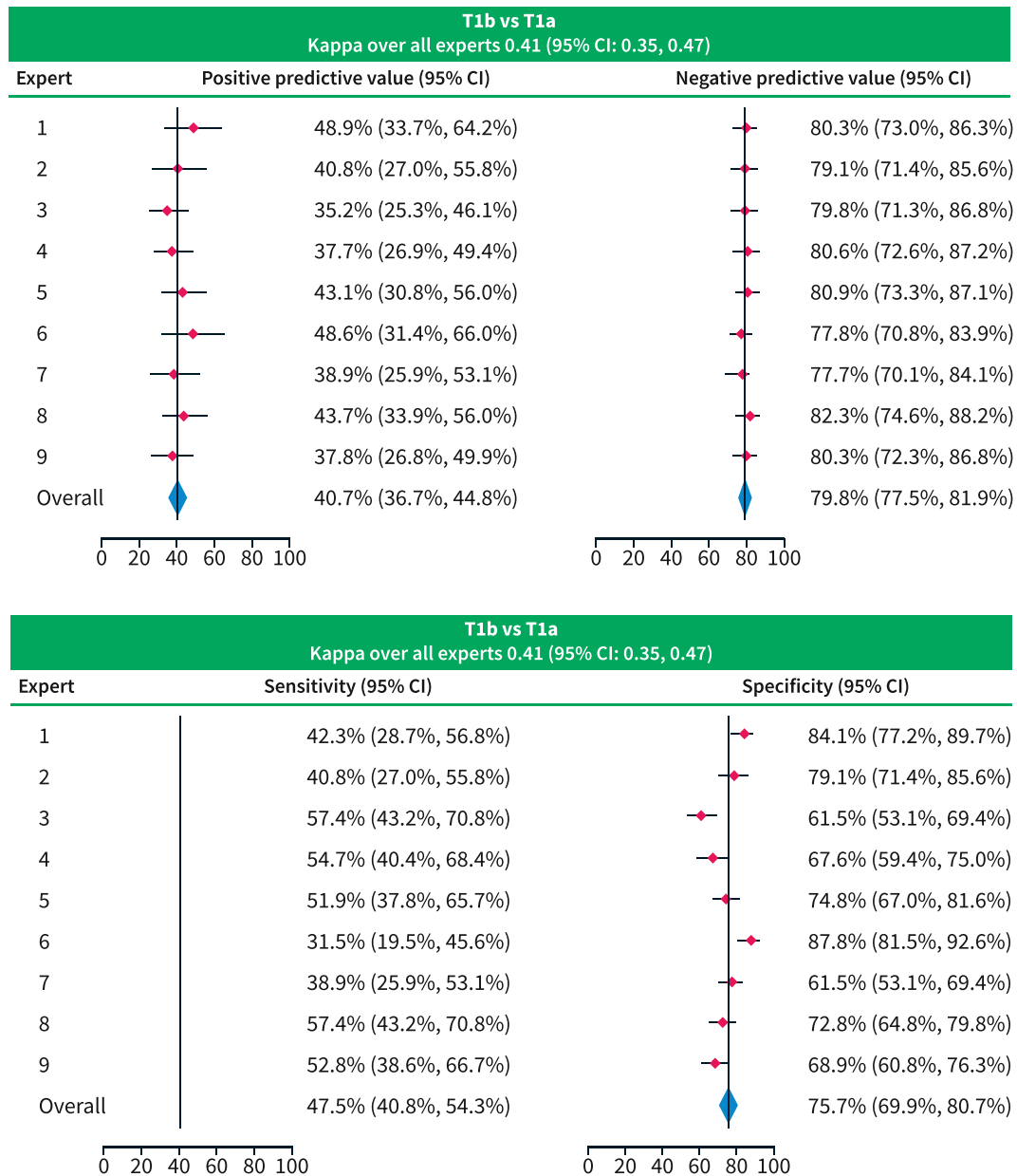| **T1b vs T1a** Kappa over all experts 0.41 (95% CI: 0.35, 0.47)		
Expert	Sensitivity (95% CI)	Specificity (95% CI)
1	42.3% (28.7%, 56.8%)	84.1% (77.2%, 89.7%)
2	40.8% (27.0%, 55.8%)	79.1% (71.4%, 85.6%)
3	57.4% (43.2%, 70.8%)	61.5% (53.1%, 69.4%)
4	54.7% (40.4%, 68.4%)	67.6% (59.4%, 75.0%)
5	51.9% (37.8%, 65.7%)	74.8% (67.0%, 81.6%)
6	31.5% (19.5%, 45.6%)	87.8% (81.5%, 92.6%)
7	38.9% (25.9%, 53.1%)	61.5% (53.1%, 69.4%)
8	57.4% (43.2%, 70.8%)	72.8% (64.8%, 79.8%)
9	52.8% (38.6%, 66.7%)	68.9% (60.8%, 76.3%)
Overall	47.5% (40.8%, 54.3%)	75.7% (69.9%, 80.7%)

FIGURE 2 Sensitivity, specificity and interobserver agreement (kappa) for the prediction of infiltration depth using two definitions/cutoffs (see text): T1b versus T1a (upper graph) and T1b ≥ sm2 versus Tbsm1/T1a.

vs. T1bsm1 plus T1a), accuracy rates of experts appeared better, ranging from 69.6% (95% CI: 62.7, 75.9) to 90.1% (95% CI: 84.9, 93.9), but kappa values were worse (overall kappa 0.27; 95% CI: 0.19, 0.36); see details in Figure 2.

Lesion assessment for Paris classification (secondary outcome)

In their assessment of lesion morphology according to the Paris classification, wide variations were seen between raters; the overall kappa value for the Paris classification was moderate with 0.51

(95% CI: 0.44, 0.60); details can be found in the Table S4. Flat and invisible lesions (Paris IIB) were not better discriminated from elevated (Paris I and IIA) and excavated lesions (Paris IIC, IIA + C, III); results were even worse, with a kappa of 0.37 (95% CI: 0.32, 0.42).

Recommendation of resection technique (secondary outcome)

Of the study patients, 110 (54.5%) were treated by ESD and 92 (45.5%) by EMR. In their recommendation whether to use ESD or EMR, experts
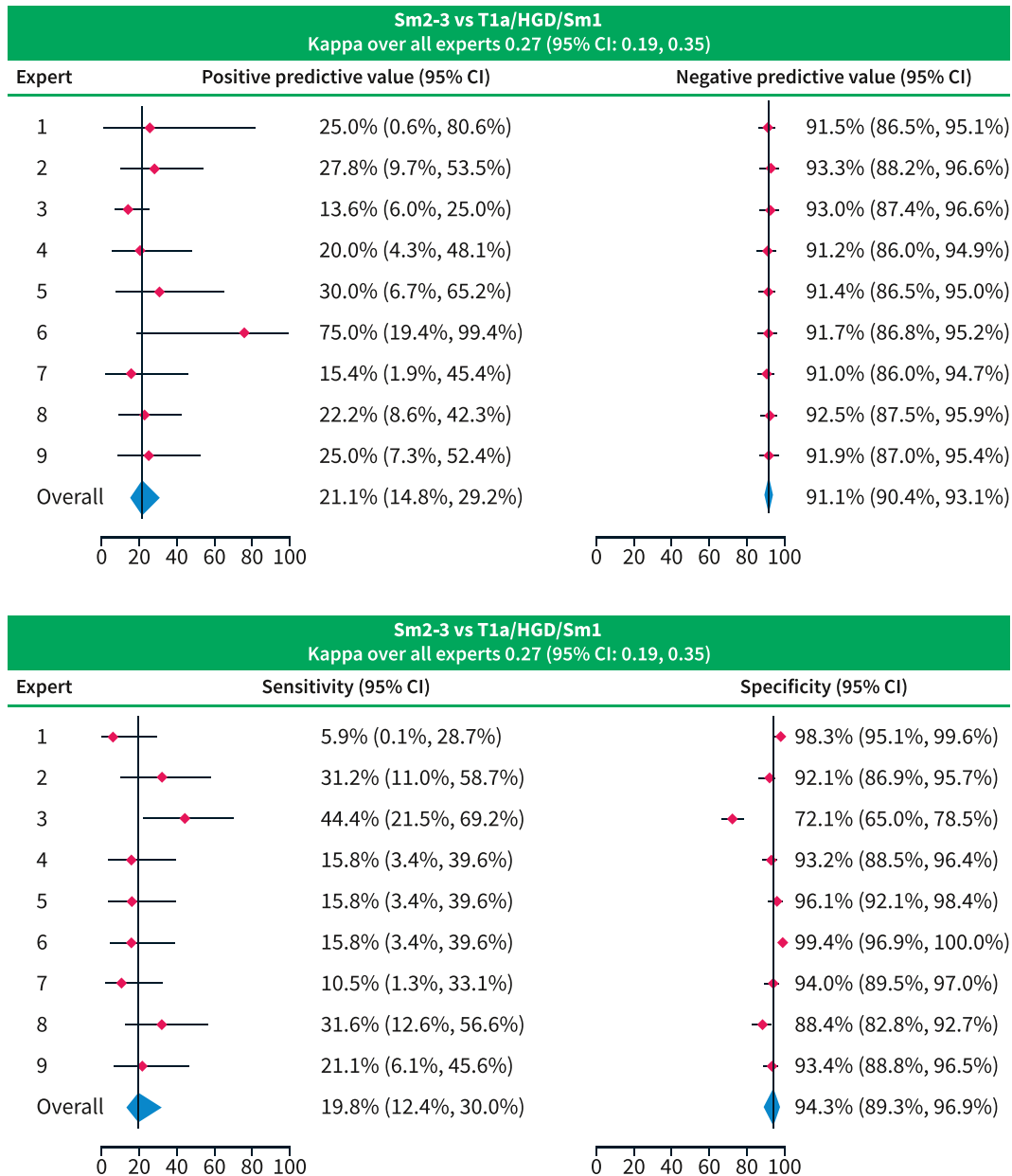
Sm2-3 vs T1a/HGD/Sm1
Kappa over all experts 0.27 (95% CI: 0.19, 0.35)

Expert	Positive predictive value (95% CI)	Negative predictive value (95% CI)
1	25.0% (0.6%, 80.6%)	91.5% (86.5%, 95.1%)
2	27.8% (9.7%, 53.5%)	93.3% (88.2%, 96.6%)
3	13.6% (6.0%, 25.0%)	93.0% (87.4%, 96.6%)
4	20.0% (4.3%, 48.1%)	91.2% (86.0%, 94.9%)
5	30.0% (6.7%, 65.2%)	91.4% (86.5%, 95.0%)
6	75.0% (19.4%, 99.4%)	91.7% (86.8%, 95.2%)
7	15.4% (1.9%, 45.4%)	91.0% (86.0%, 94.7%)
8	22.2% (8.6%, 42.3%)	92.5% (87.5%, 95.9%)
9	25.0% (7.3%, 52.4%)	91.9% (87.0%, 95.4%)
Overall	21.1% (14.8%, 29.2%)	91.1% (90.4%, 93.1%)

Sm2-3 vs T1a/HGD/Sm1
Kappa over all experts 0.27 (95% CI: 0.19, 0.35)

Expert	Sensitivity (95% CI)	Specificity (95% CI)
1	5.9% (0.1%, 28.7%)	98.3% (95.1%, 99.6%)
2	31.2% (11.0%, 58.7%)	92.1% (86.9%, 95.7%)
3	44.4% (21.5%, 69.2%)	72.1% (65.0%, 78.5%)
4	15.8% (3.4%, 39.6%)	93.2% (88.5%, 96.4%)
5	15.8% (3.4%, 39.6%)	96.1% (92.1%, 98.4%)
6	15.8% (3.4%, 39.6%)	99.4% (96.9%, 100.0%)
7	10.5% (1.3%, 33.1%)	94.0% (89.5%, 97.0%)
8	31.6% (12.6%, 56.6%)	88.4% (82.8%, 92.7%)
9	21.1% (6.1%, 45.6%)	93.4% (88.8%, 96.5%)
Overall	19.8% (12.4%, 30.0%)	94.3% (89.3%, 96.9%)

FIGURE 2 (Continued)

recommended EMR in 52.1% of all cases on average and ESD in 41.7% (Table 2). There was a wide variance in the recommended therapy.

Data shown in Table 2 (right sided columns) were also based on the assumption that the "correct" treatment allocation would be T1a = EMR and T1b = ESD; this correlation was weak with again a wide range of expert assessments. Table S5 shows the reasons why the examiners chose ESD over EMR, although they often chose not to specify their decisions; again, variability was substantial here as well: For example, the reason "infiltrating/ulcerative" was ticked between 23.5% and 70%.

Expert recommendation for primary surgery was given in a mean of 10.5 (5.2%) cases. Surgery had been performed in the actual patients in the two centers in 44 cases (21.8%),

mostly secondarily and for oncologic reasons in lesions diagnosed as high-risk histologically after removal (42; 95.5%) and only in few cases (2; 4.5%) for complications; details are shown in Table S6.

Figure 3 shows the correlation of the Paris classification and recommended resection technique, which appear to be more homogeneous. In addition, in a post hoc analysis, we wanted to know whether in certain seemingly "early" lesion morphologies such as invisible or flat ones (Paris IIB) the performance of raters was better and their recommendations more homogeneous. Therefore, we made the assumption that only lesions described as invisible or flat (Paris IIB) should lead to an EMR recommendation and could end up in a high rate of T1a histology and R0 resection. As shown in Table S7, such flat

lesions (Paris IIB) were diagnosed in 18%–33% of all cases by experts, and of those cases, EMR was recommended in more than 80% by most experts, resulting in T1a stage cancers and R0 resections again >80%.

Influence of raters

In none of the assessments including the primary and secondary study outcomes shown below, the three "internal" raters (raters 5,6,7) were in any way better or different than the remaining "external" raters (see Table 2 and Tables S2–5) Furthermore, the resection cases were done in one center, so that the internal raters could have known their own cases only.

Raters were also asked about their performance routine (% of ESD vs. EMR) in their own practice in the last 5 years. The % of ESD versus EMR did not correlate with their recommendations (Table S8 and Figure S3).

DISCUSSION

Our study shows, that, on the basis of endoscopic images, prediction of cancer infiltration depth in early Barrett cancer is difficult, if not impossible, and this reflects back negatively on the agreement on the appropriate resection technique (EMR or ESD). In every BE cancer considered possibly suitable for endoresection, there are two decisions to be made: a) whether the lesion is endoscopically fully (R0) resectable at all, and b) if so, which technique, ESD or EMR should be used. These decisions depend not only on the option to technically completely resect the lesion with free (basal) margins but also on the oncologic requirements based on final histology, possibly requiring secondary surgery or other adjuvant treatment. The latter is determined by histologic risk factors, which mostly relate to an increased

risk of lymph node metastases. Therefore, certain oncologic limitations of a "curative" endoresection exist despite of technical resection success reflected by complete/R0 resection. The superiority of one technique like ESD to provide better R0 resection is thus counterbalanced by oncologic limitations, that is, a complete (R0) resection is not necessarily an oncologic curative resection. However, some years ago, the so-called low risk criteria - G1/2 grading, mucosal infiltration, and no lymphovascular invasion, were widened in guidelines,[2,3] based on initial data by the Wiesbaden group,[17–19] followed by others.[6] More recently, low risk criteria were even further expanded by data from the Netherlands.[20] This may reflect back on the clinical utility of resection techniques, which may become even more important if the diagnosis of high-risk lesions does not necessarily lead to secondary surgery but could be treated by adjuvant therapy within multimodal protocols.
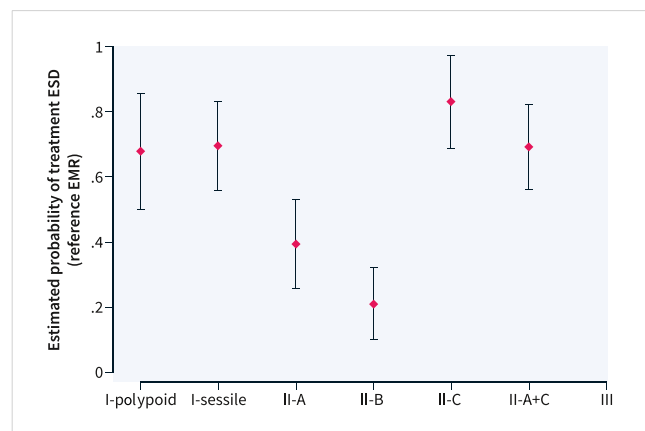

FIGURE 3 Suggested lesion treatment by endoscopic submucosal dissection (in comparison to endoscopic mucosal resection as reference) by experts in correlation with lesion morphology as assessed by the experts.

TABLE 2 Recommended resection techniques and their correlation with final *T* stage; it is assumed that for T1a lesions, EMR can be used and ESD represents a possible "overtreatment", while for tumors with submucosal infiltration, ESD should be the technique of choice due to possible deeper resection and EMR is regarded as undertreatment; surgery recommendations are not included here.

Expert	N (202)	Expert recommendations in all cases		"Adequacy" of recommendation as compared to post-resection histology in all cases			
		EMR recommendation	ESD recommendation	"Correct" treatment EMR = T1a	"Correct" treatment ESD = T1b	"Overtreatment" ESD = T1a	"Undertreatment" EMR = T1b
1	202	139 (68.8%)	63 (31.2%)	112 (80.6%)	27 (42.9%)	36 (57.1%)	27 (19.4%)
2	190	133 (70%)	49 (25.7%)	108 (81.2%)	23 (46.9%)	26 (53.1%)	25 (18.8%)
3	201	30 (14.9%)	132 (65.7%)	28 (93.3%)	34 (25.8%)	98 (74.2%)	2 (6.7%)
4	201	108 (53.7%)	85 (42.3%)	92 (85.2%)	33 (38.8%)	52 (61.2%)	16 (14.8%)
5	202	132 (65.3%)	69 (34.2%)	108 (81.8%)	29 (42%)	40 (58%)	24 (18.2%)
6	202	102 (50.5%)	96 (34.2%)	86 (84.3%)	34 (35.4%)	62 (64.6%)	16 (15.7%)
7	201	125 (62.2%)	76 (37.8%)	101 (80.8%)	30 (39.5%)	46 (60.5%)	24 (19.2%)
8	201	104 (51.7%)	71 (35.3%)	89 (85.6%)	29 (40.8%)	42 (59.2%)	15 (14.4%)
9	201	74 (36.8%)	118 (58.7%)	61 (82.4%)	34 (28.8%)	84 (71.2%)	13 (17.6%)

Note: See text for details.

For all these reasons, the decision, whether there should be an overall preference for one resection method, namely EMR (which is available more widely) or ESD (which is believed to provide better R0 resection rates, but is more difficult and risky), cannot be based on solid evidence as yet, since there is no prospective randomized trial with sufficient case numbers to provide reliable oncologic outcome data; only one small RCT looked into R0 resection rate, but was not powered for outcome.[21] Experts and guidelines therefore suggested a tailored approach, that is, that more advanced lesions should be treated by ESD, while superficial and flat lesions (as well as those with HGD only) can be treated by EMR.[2,3,6] Naturally, such a differentiated approach largely depends on full histology, which is available only after endoscopic therapy. Thus, any primary treatment allocation rests on the endoscopist's ability to predict the depth of invasion before resection; alternative techniques such as endoscopic ultrasound and CT have been shown to be mostly insufficient in these early stages.[8-10] There are only very few smaller papers on endoscopic staging ability in BE[11,22]; more data are available on early esophageal squamous cell cancer,[23] gastric[24-26] and colorectal cancer,[27-31] predominantly in the Japanese literature. In early BE cancer, ours is the first larger study to look into this issue in a systematic approach on a large dataset of images of early BE carcinoma in stages T1a (m) and T1b (sm) from 202 patients. We showed that the endoscopic aspect is not reliable for staging and subsequent treatment decisions.

Some methodological issues and limitations of our study should be discussed in detail: Our methodology was intentionally different from other interobserver studies where examiners are usually completely blinded; we added some patient information including age and sex as well as BE extent plus the results of pre-resectional biopsy. We felt that such an approach may be closer to clinical reality when experts see patients before treatment. Our procedure of looking at pre-selected images is inferior to experts' reendoscoping patients with a known diagnosis to decide on treatment and very likely also to videotapes for assessment. However, the first limitation we share with virtually all similar interobserver imaging studies published up to now.[32-37] In addition, in only a few of these studies, videos are assessed, which may also have led to different or better results. Our study imitates second opinion procedures often used in reference centers to whom images (and not videos) are sent. The exclusion of cases from a currently running randomized trial may be regarded as a further limitation, but we did not want to compromise later publication of this data. Finally, images were not taken with the aim of using them in an interobserver study, but at least to high in-house standards in one of our main areas of clinical research.

In addition, we involved four physicians (as internal raters from two centers) who had performed allocation and resection in the cases used for the study, which on the one hand may introduce a significant bias. On the other hand, these internal raters were in no way different or even better in any of the assessments (as would have been expected), which may underline the unreliability of classifications and allocations based on the endoscopic aspect. Finally, case inclusion into this study was by the performance of a resection procedure, which is naturally a further selection bias, but only this selection guarantees the reliable gold standard of histologic work-up of resection specimens. We also observed upgrading of bioptic histology—which was included in the case of CRFs the raters obtained—to more advanced forms of neoplasia after resection, a fact which is well known from the literature.[38-40]

Our results show poor predictive values and low to moderate kappa values of all experts to diagnose submucosal infiltration, and even more so to differentiate between slight (T1bsm1) and more advanced (T1b $\geq$ sm2) submucosal infiltration depths. According to changing guidelines, we chose two different cut-offs between early and more advanced T1 lesions: In the first analysis, T1sm1 (defined as submucosal infiltration $\leq$500 μm) had been included into the high-risk category (old classification), and in the second analysis, it was included in the low-risk category (new classification). In addition, complete (R0) endoscopic resection has become technically feasible in more advanced lesions and may then be followed by either risk stratification using new methods such as laparoscopic lymph node sampling[41] or by adjuvant treatment such as radiochemotherapy.[42,43] As mentioned above, expanded risk criteria of endoresection within multimodal protocols may render parameters such as R0 resection more relevant. These considerations clearly show that diagnostic tests have to be adapted to changing therapeutic standards and requirements.

Variability was lower in classifying lesions according to the macroscopic Paris classification.[44] In our study, Paris IIB was diagnosed between 20% and 30%, and Paris IIA in 30%–40% of lesions. The few papers looking into this issue were controversial, with a previous paper with all raters from the same working group reaching excellent kappa values.[45] Another Barrett paper, however, looking at detailed surface features on high resolution images reached kappa values comparable to ours.[33]

In addition, therapeutic recommendations also varied greatly between experts in our study, and did not seem to correlate with their daily practice in recent years, that is, raters with a higher ESD volume did not prefer ESD over EMR and vice versa (data not shown). On the other hand, even if the raters did not agree well about Paris lesion categories in general, so-called overtreatment (ESD in T1a, chosen in 20%–30%) could be considered clinically less relevant than undertreatment (EMR in T1b, 10.15%). Our assumption, that T1a lesions may best be treated by EMR and T1b cancers by ESD, was not mirrored by a good agreement among raters, but may not have been correct in general: Obviously, other factors such as size, lesion hardness, and lifting after injection may play an equally important role in selecting the resection technique as the Paris classification or other macroscopic consideration, similar to experience and technical proficiency of the endoscopist. All these parameters can only be assessed by a live endoscopy and not on images or videos. On the other hand, deep margins were more often negative in the ESD cases, which however cannot be directly compared to EMR

since we intentionally did not compare matched groups. Thus, further studies are necessary.

As outlined above, variability in the Paris classification was less. Thus, in invisible or flat lesions (Paris IIB), raters seemed to reach better agreements for staging and recommendation of treatment, although this is a post hoc analysis of a retrospective study and can only be considered hypothesis generating. However, we still think that an indirect conclusion from our results could be that—if expert assessment is unreliable in general - it may be better in these invisible/flat lesions with the consequences of EMR possibly sufficient in this subgroup. Should this not improve for the remaining lesions, and the macroscopic aspect not be able to guide the way, then ESD could be preferred in all other lesions due to a presumably higher oncologic efficiency with regard to complete R0 resection.

In conclusion, even experts in Barrett management could not reliably predict infiltration depth in early BE cancer stages, and the interobserver agreement was weak to moderate. This leads to deficiencies in allocating lesions to the "correct" treatment method, namely superficial (T1a) lesions to EMR, and more advanced ones (T1b) to ESD. Whether these limitations could be overcome by artificial intelligence, as some early and preliminary studies suggest,[46,47] cannot be said as yet; larger and carefully performed studies using proper histologic gold standards are required. This has to be elucidated in further clinical studies.

From a clinical management standpoint, our results could be used to guide the way to further studies on lesion allocation using either method. If only flat/invisible lesions were considered, expert agreement on the choice of therapy (EMR) was better and outcomes as well, with an R0 resection rate of more than 80%. This concerned about a quarter of cases. However, this was not the aim of our study; thus, it cannot be concluded from our data and has to be confirmed by further trials. The same is true for the assumption that all other lesions—in which expert staging assessment and interobserver agreement was poor—could then be treated with ESD as a superior method for R0 resection. Further studies are clearly necessary, starting with better definitions of lesion morphology and the need for either resection method or outcomes.

AUTHOR CONTRIBUTIONS

The study concept and performance were developed and organized by Thomas Rösch and Fadi Younis. Experts for image assessment were Torsten Beyna, Alanna Ebigbo, Siegbert Faiss, Andrea May, Oliver Pech, and Mario Anders (external) and Philip Dautel, Guido Schachschal and Thomas Rösch (internal). All authors had access to the data of the analyses and had reviewed and approved the final manuscript.

ACKNOWLEDGEMENT

CONFLICT OF INTEREST STATEMENT

None.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Fadi Younis https://orcid.org/0000-0002-5844-1795
Thomas Rösch https://orcid.org/0000-0003-2270-2495

REFERENCES

1. Sharma P. Barrett esophagus: a review. JAMA. 2022;328(7):663–71. https://doi.org/10.1001/jama.2022.13298
2. Sharma P, Shaheen NJ, Katzka D, Bergman JJ. AGA clinical practice update on endoscopic treatment of barrett's esophagus with dysplasia and/or early cancer: expert review. Gastroenterology. 2020;158(3): 760–9. https://doi.org/10.1053/j.gastro.2019.09.051
3. Pimentel-Nunes P, Libânio D, Bastiaansen BAJ, Bhandari P, Bisschops R, Bourke MJ, et al. Endoscopic submucosal dissection for superficial gastrointestinal lesions: European society of gastrointestinal endoscopy (ESGE) guideline—Update 2022. Endoscopy. 2022;54(06):591–622. https://doi.org/10.1055/a-1811-7025
4. Fotis D, Doukas M, Wijnhoven BP, Didden P, Biermann K, Bruno MJ, et al. Submucosal invasion and risk of lymph node invasion in early Barrett's cancer: potential impact of different classification systems on patient management. United Eur Gastroenterol J. 2015;3(6):505–13. https://doi.org/10.1177/2050640615581965
5. Dhingra S, Bahdi F, May SB, Othman MO. Clinicopathologic correlations of superficial esophageal adenocarcinoma in endoscopic submucosal dissection specimens. Diagn Pathol. 2021;16(1):111. https://doi.org/10.1186/s13000-021-01169-1
6. Weusten B, Bisschops R, Coron E, Dinis-Ribeiro M, Dumonceau JM, Esteban JM, et al. Endoscopic management of barrett's esophagus: European society of gastrointestinal endoscopy (ESGE) position statement. Endoscopy. 2017;49(02):191–8. https://doi.org/10.1055/s-0042-122140
7. Weusten B, Bisschops R, Dinis-Ribeiro M, di Pietro M, Pech O, Spaander MCW, et al. Diagnosis and management of Barrett esophagus: European society of gastrointestinal endoscopy (ESGE) guideline. Endoscopy. 2023;55(12):1124–46. https://doi.org/10.1055/a-2176-2440
8. Qumseya BJ, Bartel MJ, Gendy S, Bain P, Qumseya A, Wolfsen H. High rate of over-staging of Barrett's neoplasia with endoscopic ultrasound: systemic review and meta-analysis. Dig Liver Dis. 2018;50(5):438–45. https://doi.org/10.1016/j.dld.2018.02.005
9. Bartel MJ, Wallace TM, Gomez-Esquivel RD, Raimondo M, Wolfsen HC, Woodward TA, et al. Role of EUS in patients with suspected Barrett's esophagus with high-grade dysplasia or early esophageal adenocarcinoma: impact on endoscopic therapy. Gastrointest Endosc. 2017;86(2):292–8. https://doi.org/10.1016/j.gie.2016.11.016
10. Withey SJ, Goh V, Foley KG. State-of-the-art imaging in oesophagogastric cancer. Br J Radiol. 2022;95(1137):20220410. https://doi.org/10.1259/bjr.20220410
11. Ebigbo A, Mendel R, Rückert T, Schuster L, Probst A, Manzeneder J, et al. Endoscopic prediction of submucosal invasion in Barrett's cancer with the use of artificial intelligence: a pilot study. Endoscopy. 2021;53(09):878–83. https://doi.org/10.1055/a-1311-8570
12. Japan Esophageal Society. Japanese classification of esophageal cancer, 11th edition: part I. Esophagus. 2017;14:1–36. https://doi.org/10.1007/s10388-016-0551-7
13. The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: november 30 to December 1, 2002. Gastrointest Endosc 2003;58:S3–43.

14. Endoscopic Classification Review Group. Update on the paris classification of superficial neoplastic lesions in the digestive tract. Endoscopy. 2005;37(6):570–8. https://doi.org/10.1055/s-2005-861352

15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74. https://doi.org/10.2307/2529310

16. Nyaga VN, Arbyn M. Metadta: a Stata command for meta-analysis and meta-regression of diagnostic test accuracy data—a tutorial. Arch Public Health. 2022;80(1):95. https://doi.org/10.1186/s13690-021-00747-5

17. Manner H, Pech O, Heldmann Y, May A, Pohl J, Behrens A, et al. Efficacy, safety, and long-term results of endoscopic treatment for early stage adenocarcinoma of the esophagus with low-risk sm1 invasion. Clin Gastroenterol Hepatol. 2013;11(6):630–5. quiz e45. https://doi.org/10.1016/j.cgh.2012.12.040

18. Pech O, Bollschweiler E, Manner H, Leers J, Ell C, Hölscher AH. Comparison between endoscopic and surgical resection of mucosal esophageal adenocarcinoma in Barrett's esophagus at two high-volume centers. Ann Surg. 2011;254(1):67–72. https://doi.org/10.1097/sla.0b013e31821d4bf6

19. Manner H, Wetzka J, May A, Pauthner M, Pech O, Fisseler-Eckhoff A, et al. Early-stage adenocarcinoma of the esophagus with mid to deep submucosal invasion (pT1b sm2-3): the frequency of lymph-node metastasis depends on macroscopic and histological risk patterns. Dis Esophagus. 2017;30:1–11.

20. Gotink AW, van de Ven SEM, Ten Kate FJ, Nieboer D, Suzuki L, Weusten BLAM, et al. Individual risk calculator to predict lymph node metastases in patients with submucosal (T1b) esophageal adenocarcinoma: a multicenter cohort study. Endoscopy. 2021;54(02):109–17. https://doi.org/10.1055/a-1399-4989

21. Terheggen G, Horn EM, Vieth M, Gabbert H, Enderle M, Neugebauer A, et al. A randomised trial of endoscopic submucosal dissection versus endoscopic mucosal resection for early Barrett's neoplasia. Gut. 2017;66(5):783–93. https://doi.org/10.1136/gutjnl-2015-310126

22. Thomas T, Gilbert D, Kaye PV, Penman I, Aithal GP, Ragunath K. High-resolution endoscopy and endoscopic ultrasound for evaluation of early neoplasia in Barrett's esophagus. Surg Endosc. 2010;24(5):1110–6. https://doi.org/10.1007/s00464-009-0737-3

23. Inoue T, Ishihara R, Shibata T, Suzuki K, Kitagawa Y, Miyazaki T, et al. Endoscopic imaging modalities for diagnosing the invasion depth of superficial esophageal squamous cell carcinoma: a systematic review. Esophagus. 2022;19(3):375–83. https://doi.org/10.1007/s10388-022-00918-5

24. Li H, Hu H, Geng P, Guo P, Zhu Y, Zeng L, et al. The effect of short-term training about depth predicting score on the diagnostic ability of invasion depth for differentiated early gastric Cancer among non-expert endoscopists. BMC Med Educ. 2023;23(1):347. https://doi.org/10.1186/s12909-023-04230-3

25. Kang SH, Moon HS, Sung JK, Jeong H, Kim S, Kim K, et al. Endoscopic prediction of tumor invasion depth in early gastric signet ring cell carcinoma. Dig Dis. 2019;37(3):201–7. https://doi.org/10.1159/000494277

26. Nagahama T, Yao K, Imamura K, Kojima T, Ohtsu K, Chuman K, et al. Diagnostic performance of conventional endoscopy in the identification of submucosal invasion by early gastric cancer: the "non-extension sign" as a simple diagnostic marker. Gastric Cancer. 2017;20(2):304–13. https://doi.org/10.1007/s10120-016-0612-6

27. Koyama Y, Yamada M, Makiguchi ME, Sekiguchi M, Takamaru H, Sakamoto T, et al. New scoring system to distinguish deep invasive submucosal and muscularis propria colorectal cancer during colonoscopy: a development and global multicenter external validation study (e-T2 Score). Gastrointest Endosc. 2022;96(2):321–9.e2. https://doi.org/10.1016/j.gie.2022.03.002

28. Kobayashi S, Yamada M, Takamaru H, Sakamoto T, Matsuda T, Sekine S, et al. Diagnostic yield of the Japan NBI Expert Team (JNET) classification for endoscopic diagnosis of superficial colorectal neoplasms in a large-scale clinical practice database. United Eur Gastroenterol J. 2019;7:914–23. https://doi.org/10.1177/2050640619845987

29. Kawaguti FS, Franco MC, Martins BC, Segateli V, Marques CFS, Nahas CSR, et al. Role of magnification chromoendoscopy in the management of colorectal neoplastic lesions suspicious for submucosal invasion. Dis Colon Rectum. 2019;62(4):422–8. https://doi.org/10.1097/dcr.0000000000001343

30. Bisschops R, East JE, Hassan C, Hazewinkel Y, Kamiński MF, Neumann H, et al. Advanced imaging for detection and differentiation of colorectal neoplasia: European society of gastrointestinal endoscopy (ESGE) guideline - update 2019. Endoscopy. 2019;51(12):1155–79. https://doi.org/10.1055/a-1031-7657

31. Maeyama Y, Mitsuyama K, Noda T, Nagata S, Nagata T, Yoshioka S, et al. Prediction of colorectal tumor grade and invasion depth through narrow-band imaging scoring. World J Gastroenterol. 2018;24(42):4809–20. https://doi.org/10.3748/wjg.v24.i42.4809

32. Sharma P, Bergman JJ, Goda K, Kato M, Messmann H, Alsop BR, et al. Development and validation of a classification system to identify high-grade dysplasia and esophageal adenocarcinoma in barrett's esophagus using narrow-band imaging. Gastroenterology. 2016;150(3):591–8. https://doi.org/10.1053/j.gastro.2015.11.037

33. Curvers W, Baak L, Kiesslich R, Van Oijen A, Rabenstein T, Ragunath K, et al. Chromoendoscopy and narrow-band imaging compared with high-resolution magnification endoscopy in Barrett's esophagus. Gastroenterology. 2008;134(3):670–9. https://doi.org/10.1053/j.gastro.2008.01.003

34. Everson MA, Lovat LB, Graham DG, Bassett P, Magee C, Alzoubaidi D, et al. Virtual chromoendoscopy by using optical enhancement improves the detection of Barrett's esophagus-associated neoplasia. Gastrointest Endosc. 2019;89(2):247–56.e4. https://doi.org/10.1016/j.gie.2018.09.032

35. Singh M, Bansal A, Curvers WL, Kara M, Wani S, Alvarez Herrero L, et al. Observer agreement in the assessment of narrowband imaging system surface patterns in Barrett's esophagus: a multicenter study. Endoscopy. 2011;43(09):745–51. https://doi.org/10.1055/s-0030-1256631

36. Subramaniam S, Kandiah K, Schoon E, Aepli P, Hayee B, Pischel A, et al. Development and validation of the international blue light imaging for barrett's neoplasia classification. Gastrointest Endosc. 2020;91(2):310–20. https://doi.org/10.1016/j.gie.2019.09.035

37. de Groof AJ, Fockens KN, Struyvenberg MR, Pouw RE, Weusten BL, Schoon EJ, et al. Blue-light imaging and linked-color imaging improve visualization of Barrett's neoplasia by nonexpert endoscopists. Gastrointest Endosc. 2020;91(5):1050–7. https://doi.org/10.1016/j.gie.2019.12.037

38. Wani S, Abrams J, Edmundowicz SA, Gaddam S, Hovis CE, Green D, et al. Endoscopic mucosal resection results in change of histologic diagnosis in Barrett's esophagus patients with visible and flat neoplasia: a multicenter cohort study. Dig Dis Sci. 2013;58(6):1703–9. https://doi.org/10.1007/s10620-013-2689-7

39. Davis C, Fuller A, Katzka D, Wani S, Sawas T. High proportions of newly detected visible lesions and pathology grade change among patients with barrett's esophagus referred to expert centers. Dig Dis Sci. 2023;68(9):3584–95. https://doi.org/10.1007/s10620-023-07968-4

40. Yang D, King W, Aihara H, Karasik MS, Ngamruengphong S, Aadam AA, et al. Effect of endoscopic submucosal dissection on histologic diagnosis in Barrett's esophagus visible neoplasia. Gastrointest Endosc. 2022;95(4):626–33. https://doi.org/10.1016/j.gie.2021.11.046

41. Duprée A, Ehlken H, Rösch T, Lüken M, Reeh M, Werner YB, et al. Laparoscopic lymph node sampling: a new concept for patients with high-risk early esophagogastric junction cancer resected endoscopically. Gastrointest Endosc. 2021;94(2):282–90. https://doi.org/10.1016/j.gie.2021.02.014

42. Minashi K, Nihei K, Mizusawa J, Takizawa K, Yano T, Ezoe Y, et al. Efficacy of endoscopic resection and selective chemoradiotherapy for stage I esophageal squamous cell carcinoma. Gastroenterology. 2019;157(2):382–90.e3. https://doi.org/10.1053/j.gastro.2019.04.017

43. Katada C, Yokoyama T, Hirasawa D, Iizuka T, Kikuchi D, Yano T, et al. Curative management after endoscopic resection for esophageal squamous cell carcinoma invading muscularis mucosa or shallow submucosal layer - multicenter real-world survey in Japan. Am J Gastroenterol. 2022;118(7):1175–83. https://doi.org/10.14309/ajg.0000000000002106

44. Milosavljevic T, Popovic D, Zec S, Krstic M, Mijac D. Accuracy and pitfalls in the assessment of early gastrointestinal lesions. Dig Dis. 2019;37(5):364–73. https://doi.org/10.1159/000495849

45. Pech O, Gossner L, Manner H, May A, Rabenstein T, Behrens A, et al. Prospective evaluation of the macroscopic types and location of early Barrett's neoplasia in 380 lesions. Endoscopy. 2007;39(07):588–93. https://doi.org/10.1055/s-2007-966363

46. Knabe M, Welsch L, Blasberg T, Müller E, Heilani M, Bergen C, et al. Artificial intelligence-assisted staging in Barrett's carcinoma. Endoscopy. 2022;54(12):1191–7. https://doi.org/10.1055/a-1811-9407

47. Spadaccini M, Vespa E, Chandrasekar VT, Desai M, Patel HK, Maselli R, et al. Advanced imaging and artificial intelligence for Barrett's esophagus: what we should and soon will do. World J Gastroenterol. 2022;28(11):1113–22. https://doi.org/10.3748/wjg.v28.i11.1113

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.