What makes an expert Barrett's pathologist?

Concordance and pathologist expertise within a digital review panel

van der Wel, M.J.

CHAPTER 11

HISTOLOGICAL ASSESSMENT OF ENDOSCOPIC RESECTION SPECIMENS OF EARLY BARRETT'S IS HAMPERED BY POORLY DEFINED CRITERIA AND SIGNIFICANT INTER-OBSERVER VARIATION: *PROPOSITIONS TO IMPROVE ASSESSMENT*

M. J. van der Wel, E. Klaver, L. A. A. Brosens, K. Biermann, M. Doukas, C. Huysentruyt, A. Karrenbeld, F. J. W. ten Kate, G. Kats-Ugurlu, J. van der Laan, I. van Lijnschoten, F. C. P. Moll, G. J. A. Offerhaus, A. H. A. G. Ooms, C. A. Seldenrijk, M. Visser, J. G. P. Tijssen, S. L. Meijer, J. J. G. H. M. Bergman

Submitted

ABSTRACT

Background

Endoscopic resection (ER) is used to treat early cancer in Barrett's esophagus (BE). Accurate histological assessment of ER specimens is crucial for further patient management. Relevant histological features are: depth of invasion, differentiation grade, lymphovascular invasion, and basal margin radicality. We aimed to assess pathologist concordance on these histological features on ER specimens of early BE-cancer and evaluate causes of discrepancy.

Methods

62 ER cases, 3 representative slides per case were digitalized and independently assessed by 13 dedicated GI pathologists from 8 Dutch BE expert centers, using an online assessment module, annotating the four features. For each feature, concordance and discordance were measured.

Results

We observed clinically relevant discordances for assessments of all criteria. Grouping depth of invasion categories according to expanded endoscopic treatment criteria, ≥1 pathologist was discrepant in 21% of cases, increasing to 45% when grouping diagnoses according to the traditional 'mucosal vs submucosal' classification. For differentiation grade, lymphovascular invasion, and radicality of the basal margin, the discordances were substantial with 27%, 42%, and 32% of cases having ≥1 discrepant pathologist, respectively.

Conclusions

Histological assessment of ER specimens of early BE-cancer by dedicated GI pathologists shows significant discordances. We therefore propose review of ER specimens containing high-risk histological features by a second expert pathologist. Many of the observed discordances are due to poorly defined diagnostic criteria. We present propositions to attain clinically relevant subgroupings of diagnostic categories and to improve definitions of diagnostic criteria. These might be finalized into recommendations by a Delphi consensus meeting.

INTRODUCTION

Barrett's esophagus (BE) is a premalignant condition in which the stratified squamous epithelium of the distal esophagus has been replaced by columnar epithelium containing intestinal metaplasia. Patients with BE are at increased risk of esophageal adenocarcinoma (EAC). Early neoplasia in BE can be treated endoscopically by a combination of endoscopic resection and endoscopic ablation. [1-3] The cornerstone of this treatment is the endoscopic resection (ER), which effectively removes the most relevant neoplastic part of the BE and yields a resection specimen for histological examination. This histological examination provides a risk estimate for the presence of local lymph node metastases (LNM) and/or local recurrence. The most important features predictive for LNM are: submucosal invasion, poor differentiation grade, presence of lymphovascular invasion and radicality of the resection at the basal margin. [4-8] Most guidelines advocate esophagectomy for cases with one or more of these features. [9-13] Hence, a reliable diagnosis of the presence or absence of these histological features is of the utmost importance. Little is known about the observer agreement of histological assessment of ER specimens. In a recent study, two GI pathologists retrospectively assessed 25 ER specimens and were found to be discordant with the original diagnosis for depth of invasion, presence of lymphovascular invasion and tumor grade in 48%, 25% and 44% respectively. In that study, cases with a positive resection margin or interpretation difficulties due to tangential cutting were excluded. Therefore, significant discordances are likely to occur even more often. [14]

In the Netherlands, endoscopic treatment of BE neoplasia is restricted to eight BE expert centers. The majority of resection specimens obtained in these centers are assessed by a group of 13 dedicated gastro-intestinal (GI) pathologists. These pathologists constitute a national advisory panel for review of BE biopsies diagnosed as indefinite for dysplasia or low-grade dysplasia in all Dutch centers. This panel is going through a structured self-assessment program and consensus meetings to ensure homogeneity of the histological assessment of these biopsies. [15-17] We want to expand the role of our national digital review panel for BE to also include review of ER specimens. For this, participating pathologists will follow the same route of self-assessment and consensus meetings to ensure homogeneity in their histological assessment of ER specimens. The results of the first study set are presented here.

The aim of the current project was to study the histological assessment of ER specimens of early BE cancer by evaluating the diagnostic concordance of 13 BE expert pathologists and to evaluate causes of discrepancy of relevant features, in order to propose recommendations for improvement.

METHODS

Case selection and scanning

The pathology reports of the Amsterdam University Medical Center, Amsterdam (a tertiary referral center for patients with upper GI neoplasia) were queried for all ER specimens of early BE cancer obtained between 2006 and 2016. Based on the pathology report, a selection of cases was made enriched for one or more of the following features: submucosal invasion, poor differentiation grade, lymphovascular invasion, and/or irradicality of the basal margin of the specimen. Of these cases, all slides were anonymized and subsequently reviewed by the study coordinator (MW) in collaboration with an expert pathologist (SM) to select a single representative cross section. Eventually, the case set consisted of 62 cases with the following predefined number of features: 15 cases with superficial or deep submucosal invasion (m3-sm1; sm1-3); 14 cases with a poor differentiation grade (G3); 13 cases with an irradical basal resection margin and 13 cases with, or suspicious for, lymphovascular invasion. Eight cases contained multiple features. This enriched case selection was supplemented with 15 reference cases that had a maximum depth of invasion of m3. Depth of invasion was classified according to the American Joint Committee on Cancer (AJCC) guideline. [18] The medical ethical committee of the Amsterdam University Medical Center waived the need for approval for this study.

Per case, three stainings of the representative cross section (H&E, desmin (clone is 33 (MUO72-UC), titration 1:100, Biogenex, USA, Fremont, USA) and a (lympho-) endothelial marker staining (D2-40 or CD-31; clone D2-40 (M0823) and clone JC70A (M0823) respectively; Dako North America, Carpinteria, USA) were fully digitalized, using a slide scanner with a x20 microscope objective (.Slide, Olympus, Tokyo, Japan). Images were checked for focus and acuity by the study coordinator, re-scanned if necessary, and made available for viewing through the virtual slide system 'Digital Slidebox 4.5' (http://dsb.amc.nl/dsb/login.php, Slidepath, Leica Microsystems, Dublin, Ireland).

Subsequently, these virtual files were incorporated in a secure, online, custom-built histological assessment module (www.best-academia.eu).

Assessors

The assessors were 13 GI pathologists employed at one of the eight Dutch BE expert centers, who are jointly responsible for the histological assessment of all ER resection specimens of early BE neoplasia obtained in the Netherlands. They had a median ER assessment experience of 7 years (25-75% percentile: 7-15) and a median case load of 1 ER specimen per week over the preceding two years (25-75% percentile: 0.5-4).

Histological assessment criteria

Before starting the assessment round, histological assessment criteria of ER specimens were defined. Depth of invasion was scored as m1, m2, m3, sm1, sm2 or sm3 according to the AJCC guideline. [18] In cases with submucosal invasion, deepest point of invasion in relation to the muscularis mucosa was annotated and measured in micrometers. Differentiation grade was scored according to the World Health Organization (WHO) classification for tumors of the digestive system as 'well', 'moderately' or 'poorly' differentiated. [19] By this classification, a well-differentiated tumor is defined as having >95% gland formation ('G1'); a moderately differentiated tumor has 50-94% gland formation ('G2'), a poorly differentiated tumor has 0-49% gland formation ('G3'). Lymphovascular invasion was defined as 'tumor cells inside a lymphatic or blood vessel', diagnosed on HE or an immunohistochemical staining. Diagnostic possibilities were 'no'; 'yes' or 'suspicious for invasion'. For radicality of the basal margin, we defined an irradical basal margin (R1) as 'tumor touching inked basal resection margin'.

Online digital histological assessment

The pathologists received individual log-in credentials for the online histological assessment module, specifically designed for the purpose of this study. During the course of eight weeks they assessed the cases independently and in a random order. Case assessment consisted of two parts. First, the pathologists assessed the three digitalized cross sections (HE, desmin stain and an endothelial marker of a single cross section) in a dynamic way: cross sections had been digitalized to scrollable images, allowing the pathologists to zoom into areas of interest comparable to their standard microscopic assessment in daily practice. Subsequently, they documented the following information on a digital case record form: the presence or absence of

the aforementioned histological features, and the exact invasion depth (measured in micrometers) in case of submucosal invasion (see also **Appendix 1**). Second, they delineated and annotated the histological features present on a static (i.e. non-scrollable) image of the H&E slide. These annotated static images were then used in group discussions which were held after all assessments had been completed.

Group discussion and re-definition of diagnostic criteria

After completion of the assessments by all pathologists, three group discussion sessions were held in which all cases without a majority diagnosis (see below) for any of the four histological parameters were discussed with the whole group.

Outcome measurements

A pathologist's diagnosis was considered "concordant" if his/her diagnosis on a histological parameter matched that of the majority of the pathologists included in the analysis and "discordant" if his/her diagnosis differed from the majority. Concordance was expressed as the proportion of cases with unanimous agreement (13 pathologists) or a majority diagnosis (12 out of 13, 11/13 or 10/13 pathologists). Concordance was first reported considering all potential diagnostic categories as separate entities (e.g. depth of invasion in six levels), followed by grouping the diagnostic categories into clinically relevant subgroups. For depth of invasion, grouping was based on standard endoscopic treatment criteria (*i.e.* "mucosal" (m1-m2-m3) *vs.* "submucosal" (sm1-sm2-sm3)). In addition, we also grouped depth of invasion according to the expanded criteria for endoscopic treatment of early BE neoplasia (*i.e.* m1-m2-m3 plus sm1 *vs.* sm2-sm3). For differentiation grade, diagnoses of well-differentiated (G1) and moderately-differentiated (G2) cancers were combined *vs.* poorly differentiated cancers (G3). For lymphovascular invasion the diagnostic categories "yes" and "suspicious" were grouped. For radicality of the basal margin, 'not assessable' was kept separate from the R0 and R1 category.

Pre-processing analysis

Before calculating outcome measures of concordance and discordance, the homogeneity of the group of pathologists was evaluated. For this, the pathologists were classified according to the number of clinically relevant deviations. Clinically relevant deviations were defined as a discordant diagnosis for one of the aforementioned clinical subgroup diagnoses (e.g. when a single pathologist diagnosed

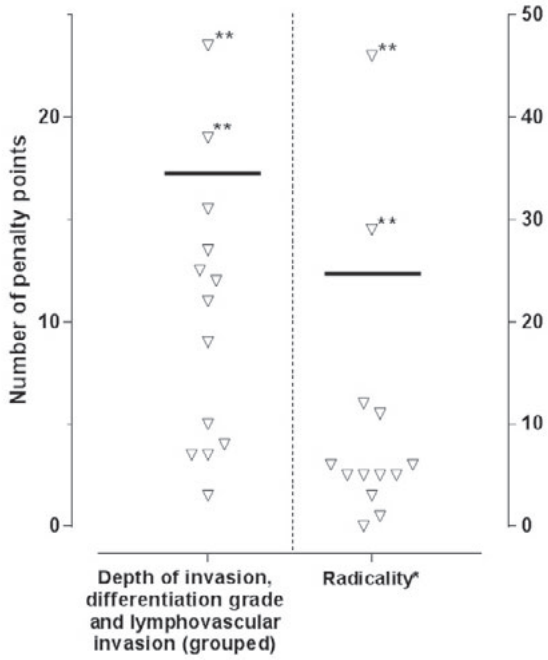
deep submucosal invasion while all other 12 pathologists diagnosed the invasion depth to be limited to the mucosa). These deviations were given "penalty points", in which three penalty points were given if a pathologist was the only one to be discordant (1 versus 12), two penalty points if the discordance was 2 versus 11, and one penalty point if the discordance was 3 versus 10. For lymphovascular invasion, a discordant diagnosis of "suspicious" was awarded with 50% of the penalty points. For radicality of the basal margin, a diagnosis 'not assessable' was also given penalty points when it deviated from the majority vote.

RESULTS

Pre-processing analysis of homogeneity

The 13 pathologists yielded >2400 slide assessments. The number of penalty points per pathologist for clinically relevant deviations was calculated to investigate the homogeneity of the group. **Figure 1** shows the penalty points of the pathologists for depth of invasion, tumor differentiation, and (lympho-)vascular invasion. The penalty points for assessing the radicality of the basal margin are depicted separately. Based on the total number of penalty points per pathologist, we arbitrarily identified four pathologists as outliers and excluded their assessments from the analyses of this study.

Figure 1: Penalty points for clinically relevant deviations compared to the majority diagnosis of the participating 13 pathologists



*Radicality depicted separately due to the diagnostic category 'not assessable'.
**denotes pathologists whose assessments were excluded from the analysis of the study.

Overall concordance

Table 1 shows the percentage of cases with different levels of concordance of nine pathologists for diagnosing depth of invasion, differentiation grade, lymphovascular invasion and radicality of the basal resection margin. Combining diagnostic categories into clinically relevant subgroups improved the concordance, yet for all four histological parameters a significant number of cases remained where one or more pathologists disagreed with the majority vote of the others.

Concordance for depth of invasion and causes of discrepancy

For depth of invasion, guideline-based subgrouping of mucosal versus submucosal cancers was associated with unanimous agreement in 34/62 cases (55%). The majority diagnosis of maximum depth of invasion was submucosal in 19 of these cases, and

mucosal in 15 cases. In 28/62 (45%) of cases at least one of the nine pathologists diagnosed a mucosal cancer while the majority diagnosed submucosal invasion, or vice versa. In 23/28 (86%) of these cases, the majority diagnosis was either m3 (11/28) or sm1 (12/28). When we used the expanded guideline criteria (i.e. grouping sm1-cancers with mucosal cancers), the number of discordant cases decreased significantly: unanimous agreement was present in 49/62 cases (79%). In the 13 cases where there was no unanimous agreement, this was due to difficulties in interpretation of the muscularis mucosae due to fragmentation by tumor in 9/13 cases (70%), due to technical artefacts in 2/13 cases (inadequate desmin staining or slide quality, 15%) and unequivocal interpretation in angle of measurement between muscularis mucosae and deepest tumor infiltration in 2/13 cases (15%). In the group discussion, the pathologists observed that the exact measurement of depth of invasion depends on three factors: the deepest point of submucosal invasion in the cross section; the interpretation of the original course of the muscularis mucosae in the area where the cancer has invaded (and inevitably partially replaced) the muscularis mucosae; and the angle of measurement between these two (**Figure 2**).

Concordance for differentiation grade and causes of discrepancy

The distinction between well-/moderately differentiated cancers (G1-G2) and poorly differentiated cancers (G3) showed unanimous agreement in 45/62 cases (73%). In the group discussions, we observed that most of the discordances were caused by differences in relating the volume of the observed poorly differentiated focus to the estimated total volume of the tumor (as required according to the WHO-criteria) and not by discordances in evaluating the architectural changes of that particular focus. This was the case in 10/17 (59%) cases without unanimous agreement. In the other 7 cases (41%), the tumor was considered a borderline G2-G3 tumor (**Figure 2**).

Concordance for (lympho-)vascular invasion and causes of discrepancy

Unanimous agreement in the diagnosis of (lympho-)vascular invasion was observed in 36/62 cases (58%). In the group discussions, this feature was considered the most difficult feature to assess. Most discordances reflected interpretational differences about whether small clusters of tumor cells were actually located within a vascular structure (20/26 cases, 76%) or whether minute foci morphologically truly consisted of tumor cells (3/26 cases, 12%). Only a minority of discordances reflected overlooked foci (3/26 cases, 12%; **Figure 2**).

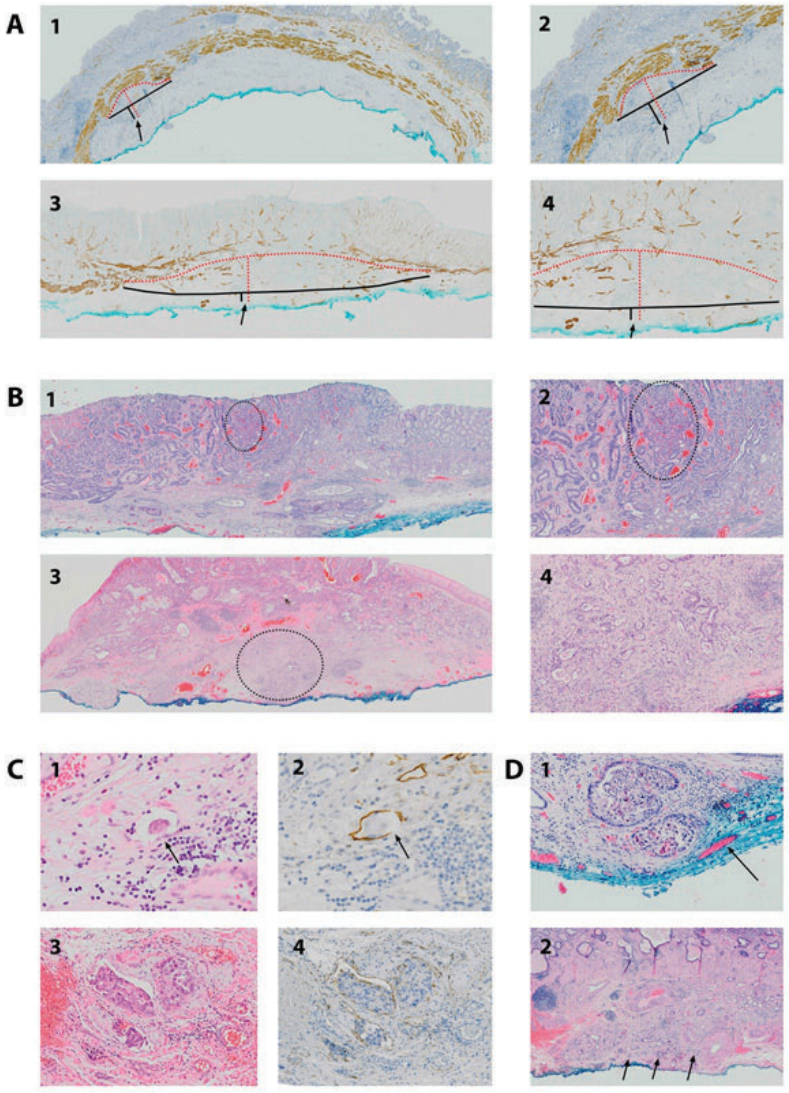
Table 1: Number and percentage of cases with different levels of concordance of nine pathologists in 62 endoscopic resection specimens of early Barrett's cancers

Feature	Number of pathologists in agreement				
Depth of invasion	9 out of 9 (%)	≥8 out of 9 (%)	≥7 out of 9 (%)	≥6 out of 9 (%)	≥5 out of 9 (%)
6 categories: m1-m2-m3-sm1-sm2-sm3	12 (19%)	26 (42%)	38 (61%)	50 (81%)	57 (92%)
"mucosal" *vs.* "submucosal"	34 (55%)	46 (74%)	54 (87%)	57 (92%)	62 (100%)
"mucosal+sm1" *vs.* "deep submucosal"	49 (79%)	51 (82%)	55 (89%)	58 (94%)	62 (100%)
Differentiation grade					
3 categories: G1-G2-G3	5 (8%)	18 (29%)	34 (55%)	47 (76%)	61 (98%)
Well/moderate differentiation vs. poor differentiation	45 (73%)	46 (74%)	54 (87%)	59 (95%)	62 (100%)
Lymphovascular invasion					
2 categories ('no' versus 'yes' and 'suspicious')	36 (58%)	47 (76%)	55 (89%)	59 (95%)	62 (100%)
Radicality					
R0 versus R1	42 (68%)	54 (87%)	58 (94%)	61 (98%)	62 (100%)

Concordance for radicality of the basal margin and causes of discrepancy

Unanimous agreement was reached in 42/62 cases (68%), indicating that clinically relevant discordances may occur in a significant number of cases where assessment of basal margin radicality is at stake. Discordances (including diagnoses of 'not assessable') were caused by differences in growth pattern interpretation in 13/20 cases, or by artefacts (including curling of the lateral margin) in 7/20 cases. In the group discussions, the discussion centered around tumor growth pattern. There were instances where the tumor growth pattern was so incohesive, that even though the tumor was actually located within 100 micrometers of the basal resection margin, but not literally 'tumor touching ink', based on the tumor architecture and close proximity, the pathologists were unable to dismiss the chance of an irradical resection.

Figure 2: Examples of variability in assessment per diagnostic category



A: Variability in measuring submucosal invasion based on different interpretation options for the continuity of the muscularis mucosa. A 1-2: Deepest measurement of invasion was 4000 micrometers (range 800-4000, red dotted line), shallowest measurement of invasion was 100 micrometers (range 100-600, black line). For this specimen, no consensus was reached for depth of invasion after discussion (originally graded as m3/sm1 by three pathologists and as sm2/sm3 by six). A 3-4: Deepest measurement of invasion was 1000 micrometers (range 600-1000, red dotted line), shallowest measurement of invasion was 40 micrometers (40-500, black line). For this specimen no consensus was reached for depth of invasion (graded as m3/sm1 by four pathologists and as sm2/sm3 by five).

B: ER specimens containing tumor with heterogenous differentiation patterns and distribution, leading to assessment discordances in grading tumor differentiation. B 1-2: A well to moderately differentiated adenocarcinoma containing a minute component of poor differentiation grade (dotted circles) in the superficial part of the tumor (originally graded as G1 by two pathologists and as G2 by seven pathologists). B 3-4: A moderately differentiated adenocarcinoma containing a small component of poor differention grade at the invasive front (arrow), final consensus diagnosis: G3 after discussion (originally graded as G2 by seven pathologists and as G3 by two pathologists).
C: Examples of lymphovascular invasion. C 1-2: A small focus of lymphovascular invasion (arrow), confirmed with vascular marker in C2 (originally diagnosed as 'not present'; by six pathologists and as 'present' by three pathologists). C 3-4: Extensive vascular invasion (arrow), confirmed with vascular marker in C4, although not circumferentially positive (originally diagnosed as 'present' by nine pathologists).
D: Examples of a radical and irradical resection. D 1: Tumor located along the invasive front, with a minor component of connective tissue and some minor vascular structures (arrow) located between the tumor and the ink, so R0 resection (originally diagnosed as R0 by three pathologists and as R1 by six).
D 2: Tumor located along the invasive front, where scattered tumor glands (arrows) are suggestive of an irradical (R1) resection. However, there is no 'tumor touching ink,' so R0 resection (originally diagnosed as R0 by five pathologists and as R1 by four).

DISCUSSION

The aim of this descriptive study was to investigate the diagnostic concordance of a group of dedicated GI pathologists on the histological evaluation of ER specimens of early BE cancer. The 13 GI pathologists participating in this study are working at the eight BE expert centers in the Netherlands, in which the care for all patients with early Barrett's neoplasia in the Netherlands is centralized. As a result, they have a high exposure to BE biopsies and ER specimens.

Clinically relevant discordances were seen in the histological diagnosis of depth of invasion, differentiation grade, lymphovascular invasion, and radicality of the basal resection margin of ER specimens of BE neoplasia. For depth of invasion, 21% of the selected specimens had 1 or more pathologist disagreeing with the majority vote of the others when categories were grouped according to the expanded criteria for endoscopic treatment (i.e. grouping sm1 with mucosal cancers). When diagnoses were grouped according to the traditional 'mucosal vs submucosal' classification, discordance was even more frequent with at least one pathologist disagreeing with the majority vote of the others in 45% of cases. Also, for differentiation grade, lymphovascular invasion, and radicality of the basal margin, the discordances were substantial with 27%, 42%, and 32% of cases having at least one discordant pathologist, respectively.

It should be noted that our set of ER specimens was purposely enriched for these histological features; the aforementioned percentages, therefore, cannot be considered to reflect the frequency of discordant diagnoses for all ER specimens of BE neoplasia. The concordance will likely be higher for those ER specimens with mucosal cancer, which constitute the majority of ER resections in BE and in which features such as poor tumor differentiation, lymphovascular invasion, and an irradical basal margin are relatively rare. Nevertheless, our study does suggest that there is a significant level of uncertainty for the "histological ground truth" of ER specimens for patients where the decision between continuing endoscopic treatment or the need to upscale to surgical resection is at stake. Therefore, we propose to seek an expert revision by at least one expert pathologist, for all endoscopic resection specimens containing a histological feature that potentially pushes the patient from endoscopic management to surgical esophagectomy or additional chemo-radiotherapy.

Our study included pathologists with a high exposure to BE neoplasia. The group consisted of all GI pathologists working at the Dutch BE expert centers. These centers manage all patients with BE neoplastia in the Netherlands according to a joint clinical protocol and the group meets twice a year for case discussions and joint training. In addition, over the last three years, all pathologists have participated in a structured self-assessment program with face-to-face group meetings, to build a national digital review panel for BE biopsy cases diagnosed 'indefinite for dysplasia' and 'low-grade dysplasia'. [15-17] In addition, we artificially improved the assessment homogeneity in the current study by excluding the four pathologists with the highest number of penalty points for clinically relevant deviations. In our opinion, our results therefore reflect the performance of expert BE pathologists. We speculate that for pathologists with less exposure to BE neoplasia, e.g. in a setting where centralized care of BE neoplasia is not implemented, variability in the assessment of clinically relevant histological features may be present at an even higher rate. After finishing the assessments, group discussions were organized to reach consensus on each feature for each case, and to evaluate causes of discrepancy. These discussions led to a series of propositions (**Table 2**) that can aid pathologists when assessing ER specimens in clinical practice.

For depth of invasion, most discordances reflected the difficulty of distinguishing m3-sm1 cancers. This distinction, however, may be clinically less relevant since patients with an sm1 cancer without G3-differentation or lymphovascular invasion are accepted

more and more as candidates for endoscopic treatment. [20 21] By grouping m3 and sm1 cancers, the presence or absence of (borderline) submucosal invasion becomes less of an issue; the focus shifts to measuring the exact depth of invasion in those cases with evident (deeper) submucosal invasion. **Figure 2** illustrates the potential variability in measuring submucosal invasion and the recommended approach (see also **Table 2**). To reduce this variability, we propose 3 measures: 1) obtain additional cuts of the deepest point of invasion to find the focus with the deepest invasion, 2) use desmin immunohistochemical staining to highlight the muscularis mucosae enabling optimal assessment of its anticipated course at the site of tumor invasion (the muscularis mucosae is generally fragmented by tumor invasion at this site), and 3) in case of unequivocal submucosal invasion, perform multiple measurements of infiltration depth in micrometers *perpendicular* to anticipated deepest margin of the invaded muscularis muosae, reporting the range of these measurements in the pathology report.

During the discussion on differentiation grade, it was questioned whether the WHO guidelines are appropriate for assessing tumor differentiation in ER specimens. Under the definition of the WHO guideline, a relatively large volume of the tumor has to have poorly differentiated features before it is classified as "poor". As a result, an early BE cancer with a 5-cm diameter intramucosal (T1a) component, containing a 0.5-cm focus of submucosal invasion with poorly differentiated features is graded as G1 (since the volume of the poorly differentiated component does not meet >50% cut-off value according to the WHO-criterium for G3), whereas a 0.5-cm cancer with a similarly sized poorly differentiated component is classified as G3. For other histological features (i.e. submucosal invasion) we refrain from using combinations of qualitative and quantitative features and instead score the most advanced component: e.g. any submucosal invasion instead of a minimum required percentage of submucosal invasion. In our opinion, it would be logical to apply the same principle to scoring the differentiation grade of ER specimens (**Table 2**). If a tumor has a heterogeneous differentiation with a poorly differentiated component, we therefore propose for the tumor to be graded as a poorly differentiated cancer.

For lymphovascular invasion, not all discordant cases were solved in the group discussions, due to differences in interpretation and certain configurations that can mimic lymphovascular invasion. Lymphovascular invasion in an ER specimen is

present in <10% of cases, since it is lacking in most mucosal cancers which constitute the majority of ER specimens. In our enriched subset, the feature was deemed present in 15/62 cases. When present, it was mostly focal and small. Logically, a presumptive diagnosis of lymphovascular invasion would be more valid if more pathologists agree on this finding and/or when it is present multi-focally instead of only a single focus. In this case set, the fact that the pathologists knew they were searching for lymphovascular invasion in an enriched dataset might have caused some overdiagnosis. Nevertheless, for the assessment of lymphovascular invasion we advise using additional cuts and IHC stainings in case of any suspicion of LVI, and to report its presence as focal (1-2 foci) or multi-focal (≥3 foci, see **Table 2**). Examples of lymphovascular invasion can be appreciated in **Figure 2**.

For the assessment of basal margin radicality, the group proposed to report the exact distance from the deepest tumor border to the basal resection margin in micrometers for all submucosal disease. Furthermore, they proposed to reserve a diagnosis of "not assessable" for those cases in which the deep margin is not completely visible because of cauterisation or other artefacts (pin prick artefact, curling of side margins, or tangential cutting). In our opinion the diagnosis "irradical" (R1) could be used only for those cases when 'tumor touching ink' is found. In this study, 2/3^{rds} of the discordant/non-assessable diagnoses was due to the interpretation of the tumor growth pattern on a single HE slide. The remaining 1/3rd of problematic interpretations was due to tissue artefacts related to endoscopic resection techniques and/or tissue processing. The main cause for disagreement in our study was the interpretation of overall tumor architecture in close proximity to the basal margin (i.e <50 micrometers), without actual 'tumor touching ink'. In such cases, the microscopic tumor growth pattern often consisted of dispersed cells within the stroma. Logically, when provided with only a single cut, this combination increases the chance for residual tumor cells to be present at the basal margin of the specimen, and therefore caused disagreements in calling those cases 'radical' or 'irradical' resections. These interpretation difficulties could be substantially decreased by providing deeper cuts and additional keratin immunohistochemical staining. This would leave only a small percentage of cases with artefacts in which the basal margin assessment remains challenging. **Table 2** summarizes the relevant recommendations for all criteria that evolved from these group discussions. Histological examples can be appreciated in **Figure 2**.

This descriptive study has a number of unique features. It is the first study to assess the diagnostic concordance on ER specimens of such a large, homogenous group of expert pathologists on such a large number of digitalized ER cross sections. This study is part of a joint training program for pathologists working at the Dutch BE expert centers. It is designed to guarantee the quality and uniformity of histological assessment for all Dutch patients treated endoscopically for early BE cancer. Our study has a number of logical limitations. The cases were preselected based on HE slides and only a single cross-section per case with a limited number of additional stainings were used for the study assessments. We aimed to make the study set uniform, and to provide IHC stainings for all cases. In some cases these IHC stainings were not routinely performed in the initial work-up and were acquired later. Therefore, these cannot strictly be interpreted as parallel sections. Second, one pathologist aided in the selection of cases and also assessed the study set. An adequate 'wash-out' period of 1 year was taken into account for this pathologist. Lastly, we are aware of the fact that we did not yet validate our propositions in **Table 2**. We are in the process of organizing a Delphi consensus meeting to further discuss these propositions and to improve the histological criteria based on international expert opinion. These criteria we will then be validated in an independent set of ER specimens.

In conclusion, the histological assessment of ER specimens shows significant variability even among expert BE pathologists. This can be partially overcome by categorizing the assessments for depth of invasion into clinically relevant groups (mucosal + sm1 versus sm2-3), which makes the assessment more simple and more reliable. For the other features, the diagnostic criteria may require further specification (see **Table 2**). In many BE expert centers, the assessment of BE neoplasia is performed by a single pathologist. To assist the aforementioned required review, a digital review platform may facilitate the exchange of digitalized microscopic images.

REFERENCES

1. Ell C, May A, Gossner L, et al. Endoscopic mucosal resection of early cancer and high-grade dysplasia in Barrett's esophagus. *Gastroenterology* 2000;118(4):670-7.

2. Peters FP, Brakenhoff KP, Curvers WL, et al. Histologic evaluation of resection specimens obtained at 293 endoscopic resections in Barrett's esophagus. *Gastrointestinal endoscopy* 2008;67(4):604-9. doi: 10.1016/j.gie.2007.08.039

3. Phoa KN, Pouw RE, Bisschops R, et al. Multimodality endoscopic eradication for neoplastic Barrett oesophagus: results of an European multicentre study (EURO-II). *Gut* 2016;65(4):555-62. doi: 10.1136/gutjnl-2015-309298

4. Stein HJ, Feith M, Bruecher BLDM, et al. Early esophageal cancer - Pattern of lymphatic spread and prognostic factors for long-term survival after surgical resection. *Annals of surgery* 2005;242(4):566-75. doi: DOI 10.1097/01.sla.0000184211.75970.85

5. Bollschweiler E, Baldus SE, Schroder W, et al. High rate of lymph-node metastasis in submucosal esophageal squamous-cell carcinomas and adenocarcinomas. *Endoscopy* 2006;38(2):149-56. doi: 10.1055/s-2006-924993

6. Barbour AP, Jones M, Brown I, et al. Risk Stratification for Early Esophageal Adenocarcinoma: Analysis of Lymphatic Spread and Prognostic Factors. *Annals of surgical oncology* 2010;17(9):2494-502. doi: DOI 10.1245/s10434-010-1025-0

7. Leggett CL, Lewis JT, Wu TT, et al. Clinical and histologic determinants of mortality for patients with Barrett's esophagus-related T1 esophageal adenocarcinoma. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association* 2015;13(4):658-64 e1-3. doi: 10.1016/j.cgh.2014.08.016

8. Davison JM, Landau MS, Luketich JD, et al. A Model Based on Pathologic Features of Superficial Esophageal Adenocarcinoma Complements Clinical Node Staging in Determining Risk of Metastasis to Lymph Nodes. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association* 2016;14(3):369-77.e3. doi: 10.1016/j.cgh.2015.10.020 [published Online First: 2015/10/31]

9. Bennett C, Vakil N, Bergman J, et al. Consensus statements for management of Barrett's dysplasia and early-stage esophageal adenocarcinoma, based on a Delphi process. *Gastroenterology* 2012;143(2):336-46. doi: 10.1053/j.gastro.2012.04.032

10. Weusten B, Bisschops R, Coron E, et al. Endoscopic management of Barrett's esophagus: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy* 2017;49(2):191-98. doi: 10.1055/s-0042-122140

11. Shaheen NJ, Falk GW, Iyer PG, et al. ACG Clinical Guideline: Diagnosis and Management of Barrett's Esophagus. *The American journal of gastroenterology* 2016;111(1):30-50. doi: 10.1038/ajg.2015.322

12. Fitzgerald RC, di Pietro M, Ragunath K, et al. British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. *Gut* 2014;63(1):7-42. doi: 10.1136/gutjnl-2013-305372

13. American Gastroenterological A, Spechler SJ, Sharma P, et al. American Gastroenterological Association medical position statement on the management of Barrett's esophagus. *Gastroenterology* 2011;140(3):1084-91. doi: 10.1053/j.gastro.2011.01.030

14. Worrell SG, Boys JA, Chandrasoma P, et al. Inter-Observer Variability in the Interpretation of Endoscopic Mucosal Resection Specimens of Esophageal Adenocarcinoma: Interpretation of ER specimens. *Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract* 2016;20(1):140-4; discussion 44-5. doi: 10.1007/s11605-015-3009-7 [published Online First: 2015/10/28]

15. van der Wel MJ, Duits LC, Pouw RE, et al. Improved diagnostic stratification of digitised Barrett's oesophagus biopsies by TP53 immunohistochemical staining. *Histopathology* 2018;72(6):1015-23. doi: 10.1111/his.13462 [published Online First: 2018 Feb 28]

16. van der Wel MJ, Duits LC, Klaver E, et al. Development of benchmark quality criteria by a panel of expert pathologists assessing whole-endoscopy Barrett's esophagus biopsy cases. *United European gastroenterology journal* 2018 doi: 10.1177/2050640618764710 [published Online First: 2018 Mar 7]

17. van der Wel MJ, Duits LC, Seldenrijk CA, et al. Digital microscopy as valid alternative to conventional microscopy for histological evaluation of Barrett's esophagus biopsies. *Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus / ISDE* 2017;30(11):1-7. doi: 10.1093/dote/dox078

18. Rice TW, Rusch VW, Ishwaran H, et al. Cancer of the esophagus and esophagogastric junction: data-driven staging for the seventh edition of the American Joint Committee on Cancer/International Union Against Cancer Cancer Staging Manuals. *Cancer* 2010;116(16):3763-73. doi: 10.1002/cncr.25146

19. Bosman FT, Carneiro, F., Hruban, R. H., Theise, N. D. . WHO classification of tumours of the digestive system. Geneva: World Health Organization 2010.

20. Kunzli HT, Henegouwen MIV, Gisbertz SS, et al. Pilot-study on the feasibility of sentinel node navigation surgery in combination with thoracolaparoscopic lymphadenectomy without esophagectomy in early esophageal adenocarcinoma patients. *Diseases of the Esophagus* 2017;30(11) doi: UNSP DOX09710.1093/dote/dox097

21. Manner H, Wetzka J, May 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. *Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus / ISDE* 2017;30(3):1-11. doi: 10.1111/dote.12462 [published Online First: 2016/03/10]

Table 2: Recommendations for assessment of endoscopic resections

Feature	Description
General	1. When a histological feature in an endoscopic resection specimen potentially pushes the patient from endoscopic management to the need for a surgical esophagectomy or additional chemo-radiotherapy, we propose expert review by at least one extra pathologist
Depth of invasion	2. In case of invasion of tumor into the submucosa: we advise to measure depth of invasion in micrometers, perpendicular from the lower margin of the muscularis mucosae to the deepest point of tumor invasion. Variability in this measurement is caused by 1) uncertainty in the identification of the deepest point of invasion of tumor cells; 2) uncertainty concerning the exact position of the lower margin of the muscularis mucosae; 3) interpretation of the angle with which these 2 points are connected. To reduce uncertainty, we propose the following measures: a. Obtain multiple additional cuts of the deepest point of invasion for optimal evaluation. b. Use anti-desmin immunohistochemical staining to highlight the course of the muscularis mucosae c. In case of destruction of the muscularis mucosae by tumor, use a virtual line representing the presumed course of the lower margin of the muscularis mucosae as starting point for the measurement of depth of invasion (**Figure 2**). d. If infiltration depth into the submucosa is not unequivocal, perform multiple measurements and report a range of uncertainty (e.g. "250-350 micrometers").
Differentiation grade	According to WHO for tumors of the digestive tract (2010): - G1 ('well'): >95% gland formation - G2 ('moderate'): 50-94% gland formation - G3 ('poor'): up to 49% gland formation 3. In case of heterogeneous differentiation grades, we propose to grade the tumour according to the poorest differentiation grade, irrespective of its relative volume.
(Lympho-) vascular invasion	4. We propose this definition: 'the unequivocal presence of tumor cells in a blood- or lymph-vessel'. a. In case of uncertainty for a single area suspicious for (lympho-) vascular invasion, obtain additional cuts and parallel immunohistochemistry using an endothelial marker with a preferentially circumferential staining result to consider the area positive. b. If possible, differentiate lymphovascular invasion from invasion in blood vessels using IHC (CD31 vs D2-40 staining). c. Report (lympho-) vascular invasion as focal (1-2 foci) or multi-focal (≥3 foci).

Feature	Description
Radicality basal margin	5. We propose a 3-tiered definition: a. If tumor infiltrates into the submucosa but does not touch the inked basal resection margin, the exact distance from the deepest tumor border to the basal resection margin should be measured (in micrometers) and mentioned in the final report. b. Obtain additional cuts and a keratin IHC stain in case this margin is less than 100 micrometers; if the growth pattern is poorly differentiated; or if substantial cauterization of the basal resection margin are present. If the additional cuts are negative for 'tumor touching ink', we propose the basal margin to be condoned free of tumor (R0). c. An irradical resection (R1) is defined as tumor invading into the inked basal resection margin ('tumor touching ink'), after evaluation of deeper cuts and performance of additional immunohistochemistry, irrespective of the growth pattern.