Greater Interobserver Agreement by Endoscopic Mucosal Resection Than Biopsy Samples in Barrett’s Dysplasia

SACHIN WANI,* SHARAD C. MATHUR,* WOUTER L. CURVERS,‡ VIKAS SINGH,* LORENZA ALVAREZ HERRERO,‡ SANDY B. HALL,* OZLEM ULUSARAC,* RACHEL CHERIAN,* DOUGLAS H. MCGREGOR,* AJAY BANSAL,* AMIT RASTOGI,* BASEM AHMED,* MANDEEP SINGH,* SRINIVAS GADDAM,* FIEBO J. TEN KATE,‡ JACQUES BERGMAN,* and PRATEEK SHARMA*

BACKGROUND & AIMS: Endoscopic mucosal resection (EMR) is an important diagnostic, staging, and therapeutic tool for patients with Barrett’s esophagus (BE)-associated neoplasia. We analyzed the histopathologic characteristics of specimens collected during EMR compared with biopsy specimens from patients with BE and assessed interobserver variability in pathologists’ assessment of EMR and biopsy specimens. METHODS: We evaluated EMR (n = 251) and biopsy (n = 269) specimens collected from patients with BE at 2 tertiary referral centers. A detailed histologic analysis was performed for each EMR and biopsy specimen to determine the grade of dysplasia, depth of the specimen, proportion of specimen with dysplasia, and quality of samples. Interobserver agreement for both biopsy and EMR specimens (among 4 experienced pathologists) was calculated by using kappa statistics. RESULTS: Histologic analysis showed that submucosa was present in the majority of EMRs, compared with biopsy specimens (88% vs 1%, P < .0001). Almost all biopsy specimens (99%) included lamina propria. However, the muscularis mucosa was observed in only 58% of biopsy specimens. For both EMR and biopsy specimens, the highest grade of dysplasia comprised ≤ 25% of the total area in > 50% of the specimens. Interobserver agreement on the diagnosis of dysplasia was significantly greater for EMR specimens than biopsy specimens (low-grade dysplasia, 0.33 vs 0.22, P < .001; high-grade dysplasia, 0.43 vs 0.35, P = .018). CONCLUSIONS: Submucosa can be examined in most samples collected from EMR; the distribution of neoplasia is focal within biopsy and EMR specimens. There is more interobserver agreement among pathologists in the analysis of EMR samples than biopsy specimens for the diagnosis of dysplasia.

Keywords: Barrett’s Esophagus; Esophageal Adenocarcinoma; Endoscopic Mucosal Resection; Dysplasia.

Barrett’s esophagus (BE) is a well-established premalignant lesion for esophageal and gastroesophageal adenocarcinoma.1 Esophageal adenocarcinoma (EAC) is one of the most rapidly rising incidence cancers in the United States, exceeding that of other cancers such as breast, colon, lung, and prostate cancer. In 2009, it was estimated that 16,470 new cases of esophageal cancer would be diagnosed in the United States, of which close to 60% would be adenocarcinomas.3 Despite all the recent advances in the diagnosis and management of this lethal cancer, the overall 5-year survival continues to be a dismal 15%–20%.

Despite the progress made in the field of molecular biomarkers, conventional histologic classification of dysplasia on routine biopsies is still the single most predictive and widely used biomarker for progression of BE patients to EAC.4 The degree of dysplasia is one of the most important determinants for surveillance intervals and management of BE patients.5 The 2 significant limitations of using dysplasia as a biomarker in the prediction of progression and management of these patients include sampling errors at the time of endoscopy and limited reproducibility (interobserver variability) among pathologists. There is a significant degree of intraobserver and interobserver variability in the interpretation of all levels of dysplasia (ie, differentiating between non-dysplastic BE [NDBE], indefinite for dysplasia [IND], and low-grade dysplasia [LGD]) in the intestinal metaplasia-dysplasia-carcinoma sequence and at the higher end (high-grade dysplasia [HGD], intramucosal and submucosal EAC), even among expert gastrointestinal pathologists.6–8

Endoscopic mucosal resection (EMR) that involves local snare excision of the neoplastic lesion has been increasingly used as both a diagnostic/staging tool and a therapeutic/curative treatment option.9,10 With the expanding endoscopic armamentarium for managing BE patients with HGD/early cancer, accurate staging is critical, and herein lies the importance of diagnostic EMRs. With EMR, the depth of tumor invasion can be established by histologic criteria that allow clinicians to treat HGD or early cancers by using endoscopic therapies with greater confidence.11 In a recent systematic review, it was shown that EMR was superior to mucosal biopsies as a diagnostic tool, resulting in a change in diagnosis in approximately 25% of the patients with HGD/early cancer (either upstaging or downstaging of the lesions).12

Limited data exist in the detailed histologic assessment of EMR specimens. In addition, the data on the reproducibility of the histopathologic findings in EMR specimens are sparse.13 Thus, the aims of this study were 2-fold: (1) to perform a

Abbreviations used in this paper: BE, Barrett’s esophagus; CI, confidence interval; EAC, esophageal adenocarcinoma; EMR, endoscopic mucosal resection; H&E, hematoxylin-eosin; HGD, high-grade dysplasia; IND, indefinite for dysplasia; LGD, low-grade dysplasia; NDBE, non-dysplastic BE.

© 2010 by the AGA Institute
1542-3565/$36.00
Material and Methods

Case Selection

Cases for this study were obtained from the Barrett’s pathology database maintained at the Academic Medical Center, Amsterdam and the Kansas City Veterans Affairs Medical Center, Kansas City. Consecutive cases were retrieved on the basis of the final submitting diagnosis from BE patients enrolled in screening and surveillance programs or endoscopic treatment protocols at these 2 tertiary referral centers from January 2000–December 2007. The set of patients for biopsy and EMR cases was not the same because the aim of the study was independent evaluation of characteristics of biopsy specimens and EMR specimens. Institutional Review Board approval was obtained at both centers.

Study Slide Set

All biopsies were obtained by using large-capacity biopsy forceps. EMR was performed at both centers by using either the endoscopic cap-assisted or “ligate and cut” technique (multiband mucosectomy). Saline injection was used in the cap-assisted technique. All tissue specimens were fixed in Bouin’s solution or 10% neutral buffered formaldehyde and then embedded in paraffin. All tissue fragments designated as biopsy specimens from one location were processed in a single tissue block. Tissue obtained during EMR was marked with India ink along its deep and side margins, serially sectioned at 2- to 3-mm intervals, and processed in multiple tissue blocks. Four-micrometer-thick sections were cut from the paraffin blocks and stained with hematoxylin-eosin (H&E).

Slides from biopsy and EMR specimens were retrieved on the basis of the final diagnosis (NDBE, LGD, HGD, or intramucosal/invasive EAC) and calculated sample size. Each biopsy specimen consisted of 1 H&E-stained slide from a single paraffin block prepared during the processing of EMR tissue. Each EMR specimen consisted of 1 H&E-stained slide from a single paraffin block prepared during the processing of EMR tissue. Thus, 1 EMR case could provide more than 1 EMR slide from a single paraffin block prepared during the processing of EMR tissue. Four-micrometer-thick sections were cut from the paraffin blocks and stained with hematoxylin-eosin (H&E). The distribution was visually categorized into 50% of the specimen).

Histologic Evaluation

A detailed histopathologic assessment of EMR specimens and its comparison to biopsy specimens were performed by a single experienced pathologist (S.M.). This involved assessment of grade of dysplasia, depth of specimen, proportion of specimen harboring neoplasia, and the quality of specimens. Dysplasia in BE was identified by standardized criteria and was recognized by the presence of cytologic and architectural abnormalities within the glandular epithelium. Each slide was classified as NDBE, IND, LGD, HGD, or EAC (intramucosal/invasive) on the basis of the revised Vienna classification; slides that did not show intestinal metaplasia were classified as “other.” The percentage of the specimen involved by the most severe lesion (LGD, HGD, or EAC) was also evaluated. The distribution was visually categorized into 4 semiquantitative categories (<10%, 11%–25%, 26%–50%, and >50% of the specimen).

The standardized criteria for each category were as follows: (1) NDBE: metaplastic columnar epithelium containing goblet cells with uniform glandular architecture, basally located nuclei with smooth membranes, minimal anisokaryosis, and preserved polarity, normal nuclear/cytoplasmic ratio; greater nuclear alterations and partial mucin depletion were acceptable when associated with evidence of inflammation, erosion, or ulceration; (2) IND: cytologic changes similar to those seen in LGD but with surface maturation or presence of inflammation; (3) LGD: glandular proliferation and crowding without complex branching, cribriform or villous architecture, hyperchromatic, enlarged nuclei with mild irregularity of nuclear membrane, and nuclear stratification extending to the surface epithelium; (4) HGD: complex cribriform or villous architecture, marked nuclear pleomorphism and irregularity of contour, increased nuclear/cytoplasmic ratio, large and irregular nuclei, full-thickness nuclear stratification, loss of polarity, and prominent mucin depletion; (5) EAC: malignant cells, singly or in groups, infiltrating beyond the basement membrane, with or without associated stromal desmoplasia; a distinction between intramucosal and submucosal invasive cancer was not performed.

The highest histologic grade for each slide was recorded. For the purposes of data analysis, specimens diagnosed as IND were combined with LGD. In addition, all biopsy and EMR specimens were examined for the deepest layer seen. The deepest identifiable layer was categorized as epithelium, lamina propria, muscularis mucosa, or submucosa. Submucosa was selected only if large submucosal vessels were identifiable deep to the muscularis mucosa (taking into account the duplication of muscularis mucosa that occurs in the setting of BE).

The quality of specimen was recorded on the basis of the following criteria: poor: extensive crush/ cautery artifact, poor orientation, and/or poor histologic processing and staining that significantly impacted definitive assessment; fair: some artifact, poor orientation, and/or poor histologic processing and staining that did not significantly impact definitive assessment; excellent: no artifact, poor orientation, and/or poor histologic processing and staining issues.

To assess the interobserver variability among pathologists in the interpretation of BE and associated neoplasia in EMR and biopsy specimens, slides were independently reviewed by 4 experienced pathologists (S.M., D.M., R.C., O.U.). All pathologists had more than 5 years of experience, with a special interest in BE. The pathologists received 2 sets of slides, 1 for biopsy specimens and 1 for EMR specimens, and were blinded to the diagnosis used for case selection.

Statistical Analysis

For the purpose of statistical analysis and on the basis of the fact that the clinical consequences and management are similar, the categories of IND and LGD were considered together. Similarly, the categories of NDBE and no intestinal
metaplasia/other were considered together. Chi-square test was used to compare the deepest layer between the groups and quality of specimen. Interobserver agreement among the pathologists was determined by using the kappa statistic, a widely used and accepted mathematical coefficient that provides a measure of interobserver agreement, accounting for agreement other than that that occurs by chance only. The strength of rater agreement was categorized according to definitions proposed by Landis and Koch for kappa values. These were as follows: 0.00–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; 0.81–1.00, almost perfect. Corresponding 95% confidence intervals (CIs) for the kappa values for each grade of histologic diagnosis were calculated. The EMR and biopsies were considered as independent samples, irrespective of whether they were obtained from the same patient and whether biopsy specimens were obtained before EMR. The kappa values for each category between the 2 groups (EMR and biopsy) were compared by using the χ² test. Statistical significance was considered at a P value of <.05. All analyses were performed by using SAS software, version 8.2 (SAS Institute Inc, Cary, NC). By following the methods proposed by Donner, it was estimated that a sample size of 220 specimens would be required for each group (EMR and biopsy) to test the hypothesis that there is an overall improvement in the kappa value from 0.4 (biopsy) to 0.6 (EMR) at α value of 0.05 and 80% power. This would ensure 55 specimens of each of the 4 categories (NDBE, LGD, HGD, and EAC) as graded by the study investigators.

Results

Detailed Histologic Analysis and Comparison of Endoscopic Mucosal Resection and Biopsy Specimens

Results of detailed histologic analysis and comparison of EMR and biopsy specimens of a single experienced pathologist are presented. A total of 251 EMR specimens (21 EAC, 54 HGD, 102 LGD, 55 NDBE, and 19 others) and 269 biopsy specimens (39 EAC, 56 HGD, 77 LGD, 89 NDBE, and 8 others) were reviewed in detail (Tables 1 and 2). The vast majority of the specimens were of excellent quality (EMR, 89%; biopsy, 97%). Across all histologic grades, the submucosa was present in the majority of EMR specimens, compared with biopsy specimens (88% vs 1.5%, P < .0001) (Figure 1). Similar results were noted for each category between the 2 groups (NDBE, 93% vs 2%; LGD, 90% vs 1%; HGD, 83% vs 0%; and EAC, 76% vs 0%; all P < .0001). An EMR specimen was less likely to show the submucosal layer than those with HGD or EAC compared with NDBE or LGD (81% vs 90%, P = .03).

Table 1. Deepest Layer Seen in EMR Specimens

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Epithelium (%)</th>
<th>Lamina propria (%)</th>
<th>Muscularis mucosa (%)</th>
<th>Submucosa (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDBE</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>4 (3)</td>
<td>51 (93)</td>
</tr>
<tr>
<td>LGD/IND</td>
<td>0 (0)</td>
<td>1 (1)</td>
<td>9 (9)</td>
<td>92 (90)</td>
</tr>
<tr>
<td>HGD</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>9 (17)</td>
<td>45 (83)</td>
</tr>
<tr>
<td>EAC</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>5 (24)</td>
<td>16 (76)</td>
</tr>
<tr>
<td>Other</td>
<td>0 (0)</td>
<td>1 (1)</td>
<td>2 (10)</td>
<td>16 (89)</td>
</tr>
</tbody>
</table>

Almost all biopsy specimens (99%) showed lamina propria (NDBE, 99%; LGD, 100%; HGD, 98%; and EAC, 95%). However, the muscularis mucosa was seen only in 58% of biopsy specimens (NDBE, 51%; LGD, 69%; HGD, 61%; and EAC, 46%) (Supplementary Figures 1 and 2). The muscularis mucosa or submucosa was seen in all cases of HGD and EAC in the EMR group, compared with 61% and 46% of specimens, respectively, in the biopsy group (P < .001). This differentiation was assessed because once cancer is beyond the muscularis mucosa and into the submucosa, the rate of nodal involvement is significantly higher compared with that of intramucosal EAC (nodal involvement of 3% or less). The percentage of the specimen involved by the most severe lesion (LGD, HGD, or EAC) is presented in Table 3. In the EMR group, a significant proportion of the specimens with neoplasia contained ≤25% of the most severe lesion (LGD, 56%; HGD, 59%; and EAC, 52%). Similarly, in the biopsy group a significant proportion of the specimens with neoplasia contained ≤25% of the most severe lesion (LGD, 50%; HGD, 28%; and EAC, 39%). Overall, a significant proportion of the specimens had the most severe lesion involving ≤25% of the entire specimen.

Interobserver Study

A total of 251 EMR and 269 biopsy specimens were reviewed by all 4 experienced pathologists. The kappa values with 95% CIs and strength of agreement for EMR and biopsy specimens in each category have been provided in Table 4. There was no difference in the overall interobserver agreement.
between the 2 groups (EMR 0.45, 95% CI [0.41–0.48] vs biopsy 0.44, 95% CI [0.41–0.47]; P = .46). Interobserver agreement was significantly superior for EMR specimens compared with biopsy specimens for the histologic diagnosis of dysplasia (LGD: 0.33 vs 0.22, P < .001; HGD: 0.43 vs 0.35, P = .018). No significant improvement in the interobserver agreement was noted for the cancer and NDBE category. Within each group (EMR and biopsy), an improvement in the level of agreement was documented from LGD to HGD to EAC.

As a subgroup analysis, the interobserver agreement was assessed by including specimens that contained >25% of the highest histologic grade. The kappa values with 95% CI and strength of agreement for EMR and biopsy specimens in each category have been provided in Table 5. With the exception of LGD, there was no statistically significant difference in the interobserver agreement between EMR and biopsy specimens. In addition, the interobserver agreement between biopsy specimens that contained ≤25% of the highest histologic grade was compared with those with >25% of the highest histologic grade (Supplementary Table 1). There was an improvement in the agreement level for all categories except LGD.

**Discussion**

Endoscopic eradication therapy is a viable therapeutic option in the management of BE with HGD and intramucosal EAC.11,13 The role of EMR as an endoscopic technique to eradicate BE-associated neoplasia continues to evolve.11,20 Recently, EMR has also been recognized as a diagnostic and staging tool. Studies suggest that EMR might be a superior diagnostic tool for grading and staging of BE-related early neoplasia compared with biopsy or endoscopic ultrasound.13,21-27

It is well-recognized that significant interobserver variability exists among pathologists in the interpretation of dysplasia in BE biopsies at the lower and higher ends of the metaplasia-dysplasia-carcinoma sequence. In a reproducibility study that involved 12 pathologists who reviewed 125 biopsy specimens, Montgomery et al8 showed that the interobserver agreement was only fair for LGD (κ = 0.32) and slight for IND (κ = 0.15), whereas for HGD/EAC the agreement was substantial (κ = 0.65) and moderate to substantial for NDBE (κ = 0.58). In another study, Downs-Kelly et al8 demonstrated poor interobserver reproducibility among pathologists in the assessment of Barrett’s mucosal biopsies at the upper end of the dysplasia spectrum (HGD, κ = 0.47; HGD with marked distortion of glandular architecture, κ = 0.21; intramucosal EAC, κ = 0.30; and submucosal invasive EAC, κ = 0.14) in 163 consecutive BE patients with at least HGD who ultimately underwent esophagectomy. Thus, given the poor interobserver reproducibility among pathologists for biopsy specimens, treatment regimens and decisions based on biopsy specimens are suspect. In a recent study, Mino-Kenudson et al13 evaluated the interobserver variability in the diagnosis of BE and associated neoplasia on EMR specimens. Nine pathologists independently reviewed mucosal biopsies and concordant EMR specimens from 25 BE patients. The intraclass correlation and the Kendall coefficient for the biopsy specimens were 0.93 (95% CI, 0.88–0.96) and 0.67, respectively, which improved to 0.97 (95% CI, 0.95–0.98) and 0.83, respectively, for EMR specimens. However, the sample size in this study was limited and not powered to assess a difference in the observer variability for each grade of dysplasia.

This large study that included evaluation of >500 EMR and biopsy specimens demonstrated that the submucosa can be examined in the majority of the EMR specimens (EMR 88% vs biopsy 1.5%, P < .0001), a critical element for the accurate diagnosis and staging of BE-associated neoplasia. On the other hand, biopsies rarely allowed assessment of the submucosa. The lower rates of submucosa in HGD/EAC compared with NDBE or LGD (81% vs 90%, P = .03) in EMR specimens could be attributed to loss of architectural landmarks as a result of fibrosis, thicker mucosa, or desmoplastic stromal reaction.

A critical distinction between HGD and EAC is the presence of infiltrating malignant cells beyond the basement membrane in EAC. Recognition of this finding requires adequate depth of tissue to visualize the architectural layers. Furthermore, the depth of invasion in EAC (intramucosal or submucosal) can only be assessed when at least the muscularis mucosa is well-represented in the specimen. Among the EMR specimens, the muscularis mucosa was seen in all cases of HGD and EAC, compared with a low 61% and 46%, respectively, in the biopsy group. These results have significant therapeutic implications because spread below the muscularis mucosa into the submucosa usually is taken as an indication for surgery as a result of the increased risk (15% or greater) of nodal involvement.

---

**Table 3. Percentage of Specimen Occupied by the Dysplastic/Cancer Lesion**

<table>
<thead>
<tr>
<th></th>
<th>1–10</th>
<th>11–25</th>
<th>26–50</th>
<th>&gt;50</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EMR</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LGD</td>
<td>43 (42)</td>
<td>14 (14)</td>
<td>31 (30)</td>
<td>14 (14)</td>
</tr>
<tr>
<td>HGD</td>
<td>21 (40)</td>
<td>10 (19)</td>
<td>13 (25)</td>
<td>8 (15)</td>
</tr>
<tr>
<td>EAC</td>
<td>8 (38)</td>
<td>3 (14)</td>
<td>3 (14)</td>
<td>7 (33)</td>
</tr>
<tr>
<td><strong>Biopsy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LGD</td>
<td>18 (24)</td>
<td>20 (26)</td>
<td>12 (16)</td>
<td>26 (34)</td>
</tr>
<tr>
<td>HGD</td>
<td>8 (14)</td>
<td>8 (14)</td>
<td>11 (20)</td>
<td>29 (52)</td>
</tr>
<tr>
<td>EAC</td>
<td>8 (21)</td>
<td>7 (18)</td>
<td>12 (31)</td>
<td>12 (31)</td>
</tr>
</tbody>
</table>

---

**Table 4. Interobserver Agreement With Strength of Agreement Among Pathologists Between EMR and Biopsy Specimens in BE**

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>EMR, κ (95% CI), strength of agreement</th>
<th>Biopsy, κ (95% CI), strength of agreement</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EAC</td>
<td>0.68 (0.63–0.73), substantial</td>
<td>0.71 (0.66–0.76), substantial</td>
<td>.3</td>
</tr>
<tr>
<td>HGD</td>
<td>0.43 (0.38–0.48), moderate</td>
<td>0.35 (0.3–0.4), fair</td>
<td>.018</td>
</tr>
<tr>
<td>LGD/IND</td>
<td>0.33 (0.28–0.39), fair</td>
<td>0.22 (0.17–0.27), fair</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>NDBE</td>
<td>0.51 (0.46–0.56), moderate</td>
<td>0.57 (0.52–0.62), moderate</td>
<td>.09</td>
</tr>
</tbody>
</table>

NOTE. Based on Landis and Koch definitions: kappa values of 0.00–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; 0.81–1.00, almost perfect.
Interestingly, this study also showed that in both groups (EMR and biopsy), a significant proportion of specimens with neoplasia occupied very focal areas (≤25% of the specimen) harboring the most severe lesion: EMR: LGD 56%, HGD 59%, and EAC 52%; and biopsy: LGD 50%, HGD 28%, EAC 39%. Because EMR specimens provide much more extensive tissue sampling, it is more likely that small foci of higher grade lesions would be identified in EMR specimens (that might be missed in smaller biopsy specimens). This might also contribute to poor interobserver agreement on biopsy specimens.

In this study that included an adequately powered sample of EMR and biopsy specimens reviewed by 4 experienced pathologists, the interobserver agreement was significantly superior for EMR specimens compared with biopsy specimens for the histologic diagnosis of dysplasia in BE (LGD and HGD). This improvement in agreement might be attributed to larger tissue sampling with EMR resulting in easier evaluation of dysplastic glands, mucosal landmarks, and improved orientation of specimens compared with superficial biopsy specimens. No significant improvement in the interobserver agreement was noted for EAC and NDBE between the 2 groups. Lack of improvement in this category might be related to changes that are obvious and easily discernable to pathologists at these extremes. When interobserver agreement was assessed for specimens (EMR and biopsy) that included >25% of the highest histologic grade, there was no statistically significant difference between the 2 groups for each histologic category with the exception of LGD. On the other hand, there was an improvement in the interobserver agreement in each histologic category, with the exception of LGD, when biopsy specimens with ≤25% and >25% of the highest histologic grade were compared.

Consistent with previous reports, the interobserver agreement for LGD in the biopsy group was low (κ = 0.22). This was true even when the biopsy specimen demonstrated LGD in >25% of the entire specimen (κ = 0.19). In addition, despite providing a much larger specimen for evaluation, the interobserver agreement for LGD in the EMR group improved but was fair (κ = 0.33). Thus, significant interobserver variability for LGD continues to be a major problem. This might partly be related to the inherent problems in the accepted definition and criteria for LGD and the overlapping histologic features with epithelial regeneration. In addition, the natural history of LGD has been well-described to be highly variable. Thus, LGD might be in need of redefinition, particularly as it relates to the risk of progression to EAC. There is a need for other biomarkers and adjunctive diagnostic techniques to improve reproducibility and help stratify cancer risk in this group of patients.

There are several limitations of this study that merit mention. The biopsy and EMR specimens were not necessarily obtained from the same site or the same patient. In addition, one EMR specimen could provide more than one section. This study was not designed to compare entire biopsy or EMR specimens, but rather to assess pathologist interobserver agreement in assessing biopsy-type (small superficial fragments) and EMR-type (large intact deeper tissue) specimens. Potential for bias as a result of these factors cannot be excluded. The method of resection and different types of electrosurgical currents used at the 2 centers might have impacted the amount of artifact present and thus the quality of EMR specimens. This study was not designed to study the relation between resection technique and size or depth of resection. All pathologists included in this study were from a single center, which might limit the external validity of the findings of this study. The detailed histologic analysis on all EMR and biopsy specimens was performed by a single experienced pathologist and not by a consensus diagnostic. However, the purpose of this endeavor was to evaluate the depth and extent of neoplasia and not the grading of neoplasia. Personal bias of the pathologist selecting both EMR and biopsy specimens cannot be excluded. This study did not assess the percentage of patients who had a change in diagnosis or outcome as a result of EMR. This was not the goal of this study, and previous reports have demonstrated that EMR results in a diagnosis change in approximately 25% of the patients with HGD/early cancer (either upstaging or downstaging of the lesions). The differentiation and interobserver agreement for intramucosal versus submucosal carcinoma was not performed for EMR specimens. Because the categories IND and LGD were considered together, analysis of the data excluding IND was not possible. This is important because IND is often used as an escape category by the pathologist, and hence IND might not necessarily fit in the intestinal metaplasia-dysplasia-carcinoma sequence. Intraobserver variability among pathologists for EMR specimens was not assessed. Assessment of the frequency of double muscularis mucosae and calculation of weighted kappa values were not performed in this study.

In conclusion, muscularis mucosa and submucosa can be examined in the majority of the EMR specimens, a critical element for the accurate diagnosis and staging of BE-associated neoplasia. Biopsies rarely allow assessment of the submucosa, and hence, treatment and management regimens for BE-related neoplasia based on biopsy specimens should be avoided. Results from this study highlight the focal distribution of neoplastic changes in both biopsy and EMR specimens. EMR significantly improves the reproducibility among pathologists for BE-related dysplasia compared with biopsy specimens. On the basis of the potential to make an accurate histologic diagnosis, EMR might represent a superior first step in the management of BE-associated neoplasia before endoscopic eradication therapies.
Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Clinical Gastroenterology and Hepatology* at www.cghjournal.org, and at doi:10.1016/j.cgh.2010.04.028.

References


Reprint requests

Address requests for reprints to: Prateek Sharma, MD, Professor of Medicine, Gastroenterology (111), Department of Veterans Affairs Medical Center, 4801 East Linwood Boulevard, Kansas City, Missouri 64128-2295. e-mail: psharma@kumc.edu; fax: (816) 922-4692.

Conflicts of interest

The authors disclose no conflicts.
Supplementary Figure 1. Biopsy specimen showing small amount of tissue with sampling of mucosal components; only the superficial muscularis mucosa is sampled (H&E; original magnification, ×5).

Supplementary Figure 2. Biopsy specimen at higher magnification showing wisps of muscularis mucosa at the base of the biopsy, likely representing the superficial layer of the split muscularis mucosa (H&E; original magnification, ×25). LP, lamina propria; MM, muscularis mucosa; SM, submucosa.
**Supplementary Table 1.** Interobserver Agreement With Strength of Agreement Among Pathologists Between Biopsy Specimens That Demonstrated ≤25% Versus >25% of Highest Histologic Grade

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Biopsy ≤25%, ( \kappa ) (95% CI), strength of agreement</th>
<th>Biopsy &gt;25%, ( \kappa ) (95% CI), strength of agreement</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EAC</td>
<td>0.66 (0.57–0.75), substantial</td>
<td>0.74 (0.68–0.80), substantial</td>
<td>.11</td>
</tr>
<tr>
<td>HGD</td>
<td>0.24 (0.15–0.33), fair</td>
<td>0.4 (0.34–0.46), fair</td>
<td>.002</td>
</tr>
<tr>
<td>LGD/IND</td>
<td>0.25 (0.16–0.33), fair</td>
<td>0.19 (0.13–0.25), slight</td>
<td>.27</td>
</tr>
<tr>
<td>NDBE</td>
<td>0.49 (0.40–0.57), moderate</td>
<td>0.59 (0.53–0.65), moderate</td>
<td>.04</td>
</tr>
</tbody>
</table>

**NOTE.** Based on Landis and Koch definitions: kappa values of 0.00–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; 0.81–1.00, almost perfect.