Endoscopic mucosal resection or ablation for Barrett's esophagus containing high grade dysplasia: agreement strongest among expert gastroenterologists

Authors Ashley Canipe[1], James Slaughter[2], Patrick Yachimski[1]

Institutions
[1] Vanderbilt University Medical Center, Division of Gastroenterology, Hepatology & Nutrition, Nashville, Tennessee 37232 United States
[2] Vanderbilt University Medical Center, Department of Biostatistics, Nashville, Tennessee 37232 United States

submitted 23. April 2014
accepted after revision 6. June 2014

Bibliography
DOI http://dx.doi.org/10.1055/s-0034-1377516
Published online: 26.9.2014
Endoscopy International Open 2014; 02: E207–E211

E-ISSN 2196-9736

Corresponding author
Ashley Canipe, MD
Vanderbilt University Medical Center
Gastroenterology, Hepatology and Nutrition
1660 The Vanderbilt Clinic
Nashville, Tennessee 37232–5280
United States
Fax: +01-615-343-8174
canipe.ashley@gmail.com

Background and study aims: Endoscopic mucosal resection (EMR) plays an important role in the staging of Barrett's esophagus (BE) and the evaluation of high grade dysplasia (HGD). The study aim is to assess the interobserver agreement among gastroenterologists expert in BE endotherapy, gastroenterologists without specified expertise in BE endotherapy, and gastroenterology trainees in recommending EMR vs ablation for BE HGD lesions, and to assess the effect of a one-time educational intervention on the interobserver agreement among non-experts and trainees.
Patients and methods: An electronic survey containing 30 still endoscopic images of BE HGD was sent to three groups of respondents: experts, non-experts, and trainees. Respondents were asked to select "Endoscopic Mucosal Resection" or "Ablation" as the most appropriate next step in management. Non-experts and trainees were then invited to repeat the survey following an educational intervention. The main outcome measure was interobserver agreement measured by Fleiss' Kappa statistic and percent agreement.
Results: In selecting between EMR and ablation, on the pre-intervention survey there was the highest amount of agreement among experts (kappa = 0.437), followed by agreement among trainees (kappa = 0.281), and non-experts (kappa = 0.107). Experts demonstrated significantly higher agreement compared to either trainees ($P < 0.001$) or non-experts ($P < 0.001$). On the post-intervention survey, interobserver agreement remained low among both trainees (kappa = 0.20) and non-experts (kappa = 0.14). Comparing the results of the surveys, there was no evidence that agreement differed for either trainees or non-experts.
Conclusions: Future efforts are needed to enable endoscopist recognition of BE HGD lesions. Consensus guidelines alone are insufficient in directing preferred endoscopic management of BE HGD.

Abbreviations

BE	Barrett's esophagus
EMR	Endoscopic mucosal resection
HGD	High grade dysplasia
NBI	Narrow band images

Introduction

Endoscopic treatment options have revolutionized therapy for Barrett's esophagus (BE) and intramucosal neoplasia. Selected patients with BE containing high grade dysplasia (HGD) and/or T1 adenocarcinoma may be candidates for endoscopic mucosal resection (EMR) and ablation rather than surgical esophagectomy. This paradigm shift has been enabled not only by the ability of endoscopic therapy to achieve disease remission but also by the ability of endoscopic diagnostic techniques to accurately stage disease.
In addition to its therapeutic role, EMR offers an important diagnostic tool by providing a robust specimen for histopathologic analysis, and is therefore frequently employed in the endoscopic staging algorithm for BE-associated neoplasia. A critical determinant is distinction between intramucosal (T1a) adenocarcinoma and adenocarcinoma invasive to the submucosa (T1b). The likelihood of lymph node involvement remains low in the case of T1a disease [1], but is considerably higher for patients with endoscopically staged T1b disease [2]. While endoscopic therapy has been reported in carefully selected cohorts of patients with T1b disease [3,4], the presence of T1b disease may prompt stronger consideration of a surgical approach to management. Numerous studies have demonstrated the ability of EMR to alter the pre-treatment diagnosis, either by upstaging to more advanced pathology or downsta-

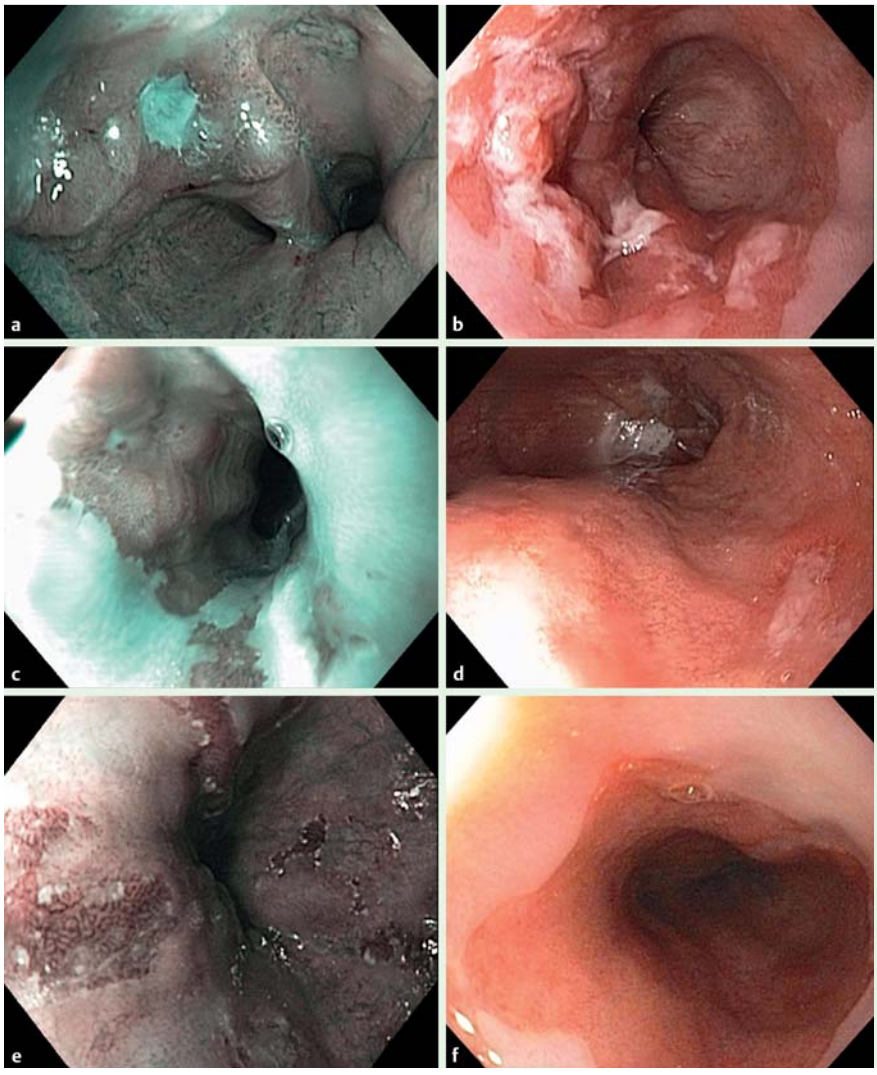

Fig. 1 Representative images from the electronic survey: **a** NBI, Paris Is lesion, **b** white light, Paris Is lesion, **c** NBI, Paris IIa lesion, **d** white light, Paris IIb lesion, **e** NBI, no visible lesion, and **f** white light, no visible lesion.

ging to less advanced pathology, among patients referred for BE endotherapy [5–7]. In recognition of the value of EMR, a recent international consensus statement recommended EMR as "essential for proper diagnosis and staging" of BE HGD associated with a visible endoscopic abnormality [8] – ideally to avoid under staging of T1a or T1b cancer prior to committing to a non-resection endoscopic ablation therapy.

Appropriate use of EMR for this purpose, independent of technical performance of EMR, first requires identification and recognition of a visible endoscopic abnormality. The ability of EMR to fulfill this role, therefore, may be influenced by factors including both an endoscopist's ability to recognize a visible endoscopic abnormality within a BE segment and awareness of consensus guidelines. Not all diagnostic examinations are performed in centers with or by endoscopists expert in diagnosis and therapy of BE neoplasia. The degree to which BE-associated neoplasia is accurately identified by non-experts or beyond expert referral centers would undoubtedly influence the clinical impact of practice guidelines pertaining to BE endotherapy.

The primary objective of this study was to examine the interobserver agreement in recommending an endoscopic diagnostic (EMR) or therapeutic (ablation) intervention among gastroenterologists expert in BE diagnosis and endotherapy, gastroenterologists without specific expertise in BE diagnosis and endotherapy, and gastroenterology trainees. A secondary objective was to assess whether a one-time educational intervention led to improved interobserver agreement among non-experts and trainees. We hypothesized that experts would demonstrate higher interobserver agreement than either non-experts or trainees in identifying lesions appropriate for EMR, and that a one-time educational intervention alone would not result in improved interobserver agreement among non-experts or trainees.

Methods

The Vanderbilt University Medical Center Institutional Review Board granted approval to conduct this study.

Endoscopic images were selected from an endoscopy database of patients who had been referred to the Vanderbilt Barrett's Esophagus Endoscopic Treatment Program (V-BEET) for endoscopic management of BE HGD. Thirty still images of lesions representing BE HGD were selected from 30 patients. These 30 images consisted of 22 high definition standard white light images and eight narrow band images (NBI). Twelve of the images demonstrated polypoid (Paris classification type 0-I) and 10 of the images represented non-polypoid (Paris classification type II-III) lesions (**Fig. 1**).

An electronic survey consisting of these 30 images was created, and respondents received an email invitation to participate in

the survey. Respondents choosing to participate then completed the survey by web link computer interface. For the pre-intervention survey, conducted in December 2012, respondents were instructed: "Indicate the most appropriate next step in management for the patient with the endoscopic abnormality shown below and a history of Barrett's esophagus with high grade dysplasia." A single endoscopic image of a Barrett's segment was shown and below the image was the option to select "Endoscopic Mucosal Resection" or "Ablation."

For the post-intervention survey, conducted in April 2013, respondents were invited to complete the same survey with images in the same order. This time, however, the survey was amended to include the following statement, viewed by respondents prior to viewing endoscopic images: "For patients with high grade dysplasia in an endoscopically visible abnormality, endoscopic resection is essential for proper diagnosis and staging". For each of the 30 images, respondents were then asked to select whether EMR or ablation would be the more appropriate next step in management, analogous to the pre-intervention survey.

Three groups of respondents were designated a priori: gastroenterologists with expertise in BE endotherapy (experts), based upon prior research or scholarly publication on BE endotherapy; attending gastroenterologists without specialized expertise in BE endotherapy (non-experts); and gastroenterology fellows (trainees). All three groups were invited to complete the pre-intervention survey. Only non-experts and trainees were invited to complete the post-intervention survey.

Assuming a correct response existed in each case, we estimated that there would be a 90% correct response rate for experts, an 80% correct response rate for non-experts, and a 70% correct response rate for trainees. A statistically significant difference would be demonstrable if the estimate of percent correct within each group was +/– 5%. The ability to achieve a 5% margin of error (half the width of the confidence interval), was therefore dependent upon two sample sizes – the number of images viewed by each respondent, and the total number of respondents in each group. With a survey content of 30 images, five experts, nine non-experts, and eleven trainees were required to achieve a 5% margin of error.

Electronic surveys were created and respondent data was stored and managed in REDCap (Research Electronic Data Capture). REDCap is a secure, web-based application hosted at Vanderbilt University and designed to support data capture for research studies [9].

Statistical analysis was performed using the R statistics program. Interobserver agreement was calculated using Fleiss' kappa statistic and also by percent agreement. Fleiss' kappa adjusts for agreement by chance alone, whereas percent agreement does not. A kappa of 0 indicates that there is no more agreement than expected by chance alone. A kappa of 1 indicates perfect agreement. Comparisons among the three groups were made by first calculating the kappa and percent agreement within a group for each of the 30 images. Paired *t* tests were then used for pairwise comparisons of the average kappa and percent agreement among the groups. A two-sided *P* value of <0.05 was considered statistically significant.

Results

Five of five invited experts responded to the survey. Nine of eleven invited non-experts responded to the survey, and fourteen of sixteen invited trainees responded to the survey. Survey respondents were asked to report professional experience data. Median number of years practicing gastroenterology (including years in training) was 11 years (IQR, 9 – 12 years) for experts, 8 (IQR, 6 – 13) for non-experts, and 2 (IQR, 1 – 2) for trainees. Median number of endoscopies performed annually for evaluation of BE was 125 (IQR, 75 – 150) for experts, 25 (IQR, 20 – 25) for non-experts, and 20 (IQR, 15 – 40) for trainees. The percent of respondents who reported performing endoscopic therapy for BE was 100% for experts, 11% for non-experts, and 0% for trainees.

Pre-intervention survey

Thirty ratings were provided by each expert, non-expert, and trainee. In selecting between EMR and ablation as the more appropriate endoscopic option, there was the highest amount of agreement among experts (kappa = 0.437). This was followed by agreement among trainees (kappa = 0.281). Non-experts had the lowest interobserver agreement (kappa = 0.107).

Reported as percent agreement, experts agreed on 87% of all responses, trainees agreed on 75% of all responses, and non-experts agreed on 67% of all responses. Experts demonstrated significantly higher percent agreement compared to either trainees ($P<0.001$) or non-experts ($P<0.001$). There was no evidence that percent agreement differed between trainees and non-experts ($P=0.49$).

Post-intervention survey

All non-experts and trainees who completed the pre-intervention survey were invited to complete the post-intervention survey. The post-intervention survey was completed by eight non-experts and ten trainees. Interobserver agreement remained low among both trainees (kappa =0.20) and non-experts (kappa =0.14).

Reported as percent agreement, trainees agreed on 74% of all responses and non-experts agreed on 77% of all responses. Expert percent agreement on the pre-intervention survey remained higher than the post-intervention percent agreement for either trainees ($P<0.001$) or non-experts ($P<0.001$). There was no evidence that percent agreement differed between trainees and non-experts ($P=0.28$). Comparing the results of the pre-intervention and post-intervention surveys, there was no evidence that percent agreement differed for either trainees ($P=0.63$) or non-experts ($P=0.08$), although non-experts demonstrated a trend toward higher agreement.

Comparison to a gold standard

It is often useful to compare individual raters to the actual, or gold standard, recommendation. In this survey, the consensus recommendation among experts (the recommendation selected by three or more of the experts) was chosen as the gold standard. Because the gold standard is defined using the consensus of experts, a higher kappa statistic and percent agreement are expected when comparing the experts to the gold standard.

On the pre-intervention survey, median kappa values relative to the gold standard were 0.70 for experts, 0.42 for trainees, and 0.20 for non-experts (**Fig. 2**). Median percent agreements were 87% for experts, 73% for trainees, and 70% for non-experts.

On the post-intervention survey, median kappa values relative to

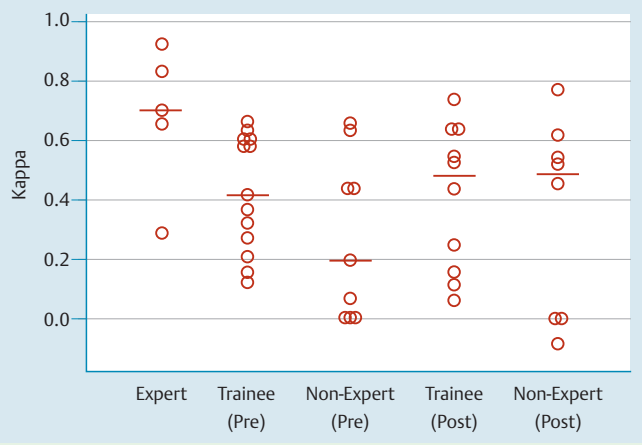

Fig. 2 Plot of kappa values relative to the gold standard (expert) on both the pre-intervention (pre) and post-intervention (post) surveys for non-experts and trainees. Each circle represents an individual rater and the horizontal bars represent the median kappa value for each group.

the gold standard were 0.48 for trainees and 0.49 for non-experts (**Fig. 2**). Median percent agreements were 75 % for trainees and 77 % for non-experts. Comparing the results of the pre- and post-intervention study, there was no evidence that, relative to the gold standard, kappa or percent agreement changed for trainees (P=0.88 and P=0.93, respectively) or non-experts (P=0.70 and P=0.38, respectively). EMR was offered as a response to all 30 questions in both surveys by three non-experts; kappa value for these three non-experts was therefore zero (**Fig. 2**).

Discussion

EMR is currently recommended for the accurate diagnosis and staging of BE HGD associated with a visible endoscopic abnormality. In this study, gastroenterologists expert in BE endotherapy demonstrated higher interobserver agreement than either non-expert or trainee gastroenterologists in selecting between EMR vs ablation for BE HGD lesions. The agreement of non-expert and trainee gastroenterologists did not improve or approach the level of experts following an educational intervention highlighting the role of EMR. The modest interobserver agreement achieved by experts (kappa=0.437), however, underscores the inherent variability and lack of a true gold standard in subjective interpretation of BE lesions and selection of endoscopic therapeutic modality.

In comparison to other endoscopic mucosal therapeutic modalities, EMR adds diagnostic value in that it offers a simultaneous diagnostic specimen of the target lesion. Accurate endoscopic staging of BE HGD is critical prior to committing patients to endoscopic or surgical intervention. For patients down staged from HGD to low grade dysplasia, radiofrequency ablation may still be offered [10], although this practice is not universal. Given low reported progression rates from low grade dysplasia to the combined endpoint of HGD/cancer [11], surveillance alone is an option for such patients. Alternatively, under staging of superficially invasive adenocarcinoma prior to proceeding with ablation therapy is undesirable. While uncontrolled retrospective data suggest that carefully selected patients with T1b disease may achieve reasonably durable outcomes following endoscopic therapy [3, 4], and that overall survival is influenced by the presence of lymph node involvement rather than a surgical intervention [4], a commitment to endoscopic therapy with a stated goal of curative intent in such individuals is understandably complex and without guarantee.

The study design, with a pre-intervention and post-intervention survey of non-experts and trainees, was intended to assess whether promotion of existing guidelines ("For patients with high grade dysplasia in an endoscopically visible abnormality, endoscopic resection is essential for proper diagnosis and staging" [8]) alone would be sufficient to bring agreement with non-experts and trainees on par with experts. Had there been improvement in agreement among non-experts and trainees following this intervention, one might conclude that improved awareness and education surrounding existing practice guidelines would be adequate for optimal endoscopic management of BE. Rather, consistent with the study hypothesis, agreement among non-experts and trainees did not improve following an educational intervention.

Given the importance of accurate pre-treatment staging, EMR should always be the initial endoscopic intervention in patients with BE-associated neoplasia. Appropriate application of EMR in this context requires two distinct sets of skills: 1) the ability to recognize a lesion; 2) technical performance of EMR. This study was designed to assess the former. Acquisition of skills necessary for accurate lesion recognition is a necessary prerequisite for delivery of appropriate therapy. With the use of high definition white light endoscopy and advanced imaging techniques including NBI, some would argue that BE-associated neoplasia is nearly always associated with a visible mucosal abnormality. The difference in agreement between experts and non-experts/trainees in this study may reflect suboptimal training and/or experience in lesion recognition among non-experts/trainees, and this difference was not remediated by a simple prompt designed to promote awareness of and adherence to treatment guidelines.

The study intervention was a weak one that consisted of a brief statement of current practice guidelines. The study was designed deliberately to highlight the fact that guidelines are not a substitute for education and training, and to underscore the notion that publication and dissemination of consensus practice guidelines are insufficient to alter clinical practice. Future educational efforts and/or experiences designed to enable endoscopist visual evaluation of BE HGD are warranted, and may serve as the basis for future study. A variation on this study with the educational intervention consisting of a didactic video on the subject of endoscopic BE lesion recognition may be an appropriate next step in this regard.

Further limitations of this study include the relatively small number of survey items and survey respondents, although both samples sizes were derived based on a priori statistical calculations. The participation rate among non-experts and trainees in the post-intervention survey was less than 100 % and created the potential for bias. All non-expert and trainee respondents were recruited from academic medical centers and the study did not assess or control for training regarding BE diagnosis and endotherapy received by these individuals prior to study participation, for instance, as in a fellowship didactic or continuing medical education settings. Moreover, the study did not assess the rationale that led respondents to choose EMR or ablation. Understanding why respondents choose a particular therapeutic modality could lead to future targeted interventions to improve interobserver variability.

In an effort to expose respondents to a range of appearances and modalities for lesion recognition, the survey contained both white and light and NBI images, and both nodular and flat BE lesions. The study did not assess familiarity with NBI use or interpretation of NBI images. A survey consisting of only white light images or only NBI images (or both for each representative lesion) may have offered a more comprehensive opportunity to assess lesion recognition skills among survey respondents. An additional alternative would have been to offer a survey consisting of endoscopic videos, rather than still images, for detailed evaluation of lesions in a format representative of real-time endoscopic practice. However, there would be challenges in creating a video that would not generate bias by "prompting" or "directing" a viewer to a focal abnormality given extra attention during endoscopic inspection. Moreover, any improvement in interobserver agreement offered by videos would likely be evident not only for non-experts and trainees but also for experts.

Potential implications of these study findings pertain to the future dissemination of BE endotherapy beyond expert referral centers. An international panel offered a strong consensus for the statement that endoscopic therapy for HGD/T1a cancer should be performed in referral centers with appropriately trained endoscopists and pathologists [8]. While data from a multicenter registry suggest that radiofrequency ablation can be performed in a community setting with acceptable efficacy and safety outcomes, most of this cohort consisted of patients with BE without dysplasia [12]. Based on the results of a recent survey of BE practice patterns, only a minority (13 %) of community-based gastroenterologists report performing EMR at this time [13].

Should use of EMR become more prevalent beyond referral centers, the findings of our study imply that education and training efforts should include development of skills in BE lesion recognition in addition to skills in the technical performance of EMR. Given the finding that recognition by non-experts of what constitutes a visible endoscopic abnormality in BE HGD may be suboptimal, the study results further suggest that patients with BE neoplasia not thought to have a visible abnormality should be referred to centers expert in BE diagnosis and therapy whenever feasible.

In summary, endoscopists with expertise in BE endotherapy have a significantly higher degree of agreement when selecting between EMR and ablation for BE HGD compared with either non-experts or trainees. The agreement of non-experts and trainees did not improve following an educational intervention, consisting of a statement endorsing the role of EMR for staging of BE HGD. Based on these results, we conclude that future efforts are needed to enable endoscopist recognition of BE HGD lesions requiring EMR, and that consensus guidelines alone are insufficient in directing preferred endoscopic management of BE HGD.

Competing interests: None

References

1 *Dunbar KB*, *Spechler SJ*. The risk of lymph node metastases in patients with high grade dysplasia or intramucosal carcinoma in Barrett's esophagus: a systematic review. Am J Gastroenterol 2012; 107: 850 – 863
2 *Prasad GA*, *Buttar NS*, *Wongkeesong LM* et al. Significance of neoplastic involvement of margins obtained by endoscopic mucosal resection in Barrett's esophagus. Am J Gastroenterol 2007; 102: 2380 – 2386
3 *Manner H*, *Pech O*, *Heldmann Y* 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: 630 – 635
4 *Tian J*, *Prasad GA*, *Lutzke LS* et al. Outcomes of T1b esophageal adenocarcinoma patients. Gastrointest Endosc 2011; 74: 1201 – 1206
5 *Moss A*, *Bourke MJ*, *Hourigna LF* et al. Endoscopic resection for Barrett's high grade dysplasia and early esophageal adenocarcinoma: an essential staging procedure with long-term therapeutic benefit. Am J Gastroenterol 2010; 105: 1276 – 1283
6 *Wani S*, *Abrams J*, *Edmundowicz SA* 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: 1703 – 1709
7 *Ayers K*, *Shi C*, *Washington K* et al. Expert pathology review and endoscopic mucosal resection alters the diagnosis of patients referred to undergo therapy for Barrett's esophagus. Surg Endosc 2013; 8: 2836 – 2840
8 *Bennett C*, *Vakil N*, *Bergman J* et al. Consensus statement for management of Barrett's dysplasia and early-stage esophageal adenocarcinoma, based on a Delphi process. Gastroenterology 2012; 143: 336 – 346
9 *Harris PA*, *Taylor R*, *Thielke R* et al. Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009; 42: 377 – 381
10 *Spechler SJ*, *Sharma P* et al. American Gastroenterological Association medical position statement on the management of Barrett's esophagus. Gastroenterology 2011; 140: 1084 – 1091
11 *Wani S*, *Falk GW*, *Post Y* et al. Risk factors for progression of low-grade dysplasia in patients with Barrett's esophagus. Gastroenterology 2011; 141: 1179 – 1186
12 *Lyday WD*, *Corbett FS*, *Kuperman DA* et al. Radiofrequency ablation of Barrett's esophagus: outcomes of 429 patients from a multicenter community practice registry. Endoscopy 2010; 42: 272 – 278
13 *Singh M*, *Gupta N*, *Gaddam S* et al. Practice patterns among U.S. gastroenterologists regarding endoscopic management of Barrett's esophagus. Gastrointest Endosc 2013; 78: 689 – 695