Open Access

Observer Variability in Gastric Neoplasm Assessment Using the Vessel Plus Surface Classification for Magnifying Endoscopy with Narrow Band Imaging

Chan Hui Yoo, Moo In Park, Seun Ja Park, Won Moon, Hyung Hun Kim, Jun Young Song and Do Hyun Kim

Department of Internal Medicine, Kosin University College of Medicine, Busan, Korea

Background/Aims: Recent studies have demonstrated that magnifying endoscopy with narrow band imaging (ME-NBI) facilitates differentiation of early gastric cancer from gastric adenoma using vessel plus surface (VS) classification. This study estimated the interobserver and intraobserver agreement of endoscopists using the Yao VS classification system for the gastric mucosal surface.

Methods: We retrospectively reviewed patients who underwent endoscopic submucosal dissection or endoscopic mucosal resection, and selected cases in which preoperative ME-NBI was conducted. Before testing endoscopists, a 20-minute training module was given. Static ME-NBI images ($n=47$ cases) were presented to seven endoscopists (two experts and five trainees) who were asked to assess the images in 20 seconds using the Yao VS classification system. After 2 weeks, the endoscopists were asked to analyze the images again. The κ statistic was calculated for intraobserver and interobserver variability.

Results: The mean κ for intraobserver agreement was 0.69 (experts, 0.74; trainees, 0.64). The mean κ for interobserver agreement was 0.42 (experts, 0.49; trainees, 0.40).

Conclusions: We obtained reliable results as assessed by observer variability, with only brief training on VS classification. The VS classification appears to provide an objective assessment of ME-NBI for trainees who are not familiar with ME-NBI.

Key Words: Endoscopy; Neoplasms; Stomach

INTRODUCTION

The ability to distinguish between benign and malignant lesions by endoscopy is critical. However, endoscopy using conventional white light imaging (C-WLI) alone is insufficient for accurate diagnosis.[1] Histological findings after resection sometimes show that low-grade adenoma has transitioned to high-grade adenoma or early cancer. Pathologists have difficulty making an accurate diagnosis using only a small biopsy specimen, and endoscopists are challenged to obtain a target biopsy from the most suspicious part of the lesion using C-WLI endoscopy alone. Differentiating low-grade adenoma, high-grade adenoma, and early cancer is therefore difficult, even after histological biopsy-based examination.[2]

Magnifying endoscopy with narrow band imaging (ME-NBI) is a powerful tool for diagnosing superficial neoplasms in the gastrointestinal tract. The NBI system is an endoscopic imaging technique for the enhanced visualization of microvascular (MV) architecture and microsurface (MS) structures of the superficial part of the mucosa.[3] Combining ME and an NBI system allows simple and clear visualization of MS structures and MV patterns of the superficial mucosa, which could be useful for obtaining a precise endoscopic diagnosis that matches the histopathological diagnosis.

Several studies have shown the potential utility of micro-

scopic capillaries, which can be visualized by ME-NBI, for predicting gastric neoplasia among superficially depressed, flat, or elevated, early gastric neoplastic lesions.[4-6]

Endoscopic diagnosis is subjective and often shows relatively low interobserver concordance. In a review of the ME-NBI literature, we found few studies on observer agreement, with most studies involving expert endoscopists. This study aimed to estimate interobserver and intraobserver variability based on the vessel plus surface (VS) classification of Yao et al.[7]

MATERIALS AND METHODS

Patients and ME-NBI images

We retrospectively evaluated our endoscopic submucosal dissection (ESD) or endoscopic mucosal resection (EMR) databases for patients treated at the Kosin University Gospel Hospital, Busan, Korea from January 2011 through June 2012. Endoscopic examinations, ESD (n=99), and EMR (n=199) were performed by two expert endoscopists. Pretreatment ME-NBI endoscopic examinations were performed before 32 ESDs and 18 EMRs. A GIF-H260Z endoscope (CLV 260 SL; Olympus, Tokyo, Japan) was used for preoperative endoscopic examinations. Pathological diagnostic criteria were based on the revised Vienna classification: four (noninvasive high-grade neoplasia) or five (invasive neoplasia) lesions were regarded as carcinoma, whereas three (noninvasive low-grade neoplasia) lesions were regarded as adenoma.[8]

We analyzed both depressed and nondepressed lesions. We excluded one case in which the image was unfocused, making discrimination difficult. The images in two cases were used as training images (Figs. 1, 2). To minimize selection bias, ME-NBI images were selected by an endoscopist who did not perform the endoscopic examinations and who was blinded to the clinical information. In total, 47 cases were selected. Two static ME-NBI images per case were selected. We selected only ME-NBI images without white light endoscopy imaging to avoid affecting the judgment of the endoscopists.

Education and review of investigators

Seven endoscopists were divided into two groups: two experienced endoscopists who had used ME-NBI occasionally for >3 years, and five trainees who had been performing endoscopies for <1 year and had never used ME-NBI. We first explained the VS classification system to the endoscopists using the image from the original publication by Yao et al.[7] and static images from two of our cases.

ME-NBI diagnostic criteria were based on the VS classification system proposed by Yao et al.,[7] which describes lesions as having 1) an irregular MV pattern with a demarcation line between the lesion and the surrounding area, and 2) an irregular
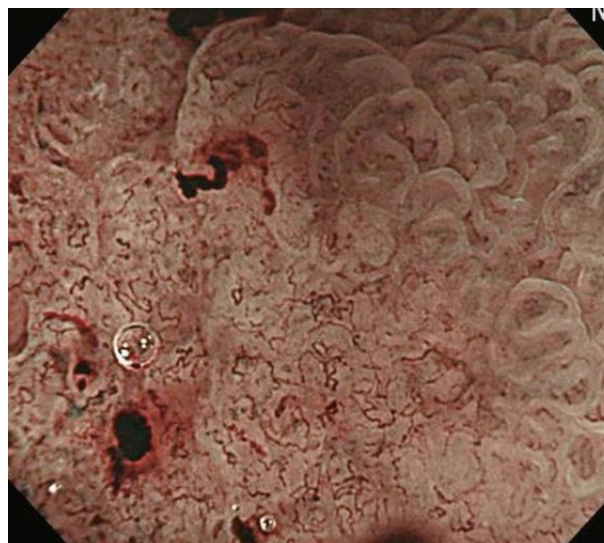
Fig. 1. Magnifying endoscopy with narrow band imaging showing a typical finding of demarcation line and irregular microvascular pattern.
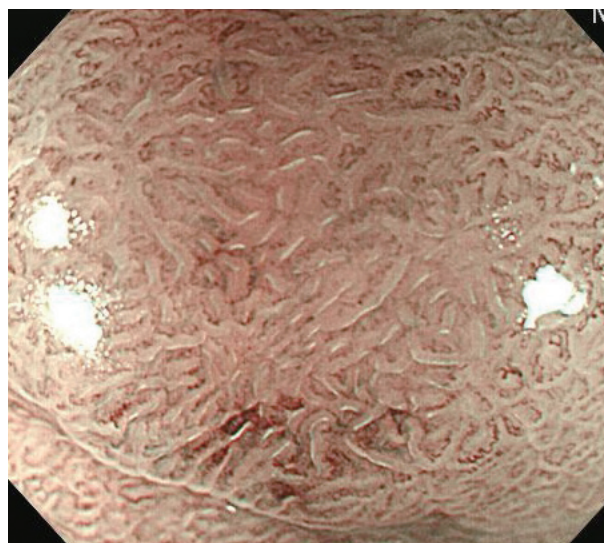
Fig. 2. Magnifying endoscopy with narrow band imaging showing a typical finding of regular microvascular and microsurface patterns

MS pattern with a demarcation line between the lesion and the surrounding area. If a target lesion has a finding of 1) or 2), it is considered positive.

We did not provide any information that might have affected the endoscopists' assessments. The seven endoscopists assessed the VS pattern of 47 cases (two images per case) for 20 seconds per case. Two weeks later, under the same conditions, re-evaluation without retraining was performed to calculate intraobserver agreement.

Statistical analysis

We used chi-square and Fisher exact tests for categorical

comparison of the data. A $p<0.05$ was considered statistically significant. Intraobserver and interobserver accordance was evaluated by the κ value, which varies from -1 (complete disagreement) to +1 (complete agreement), with zero indicating agreement by chance only. κ Values >0.8 indicated almost perfect agreement, 0.6 to 0.79 indicated substantial agreement, 0.4 to 0.59 indicated moderate agreement, 0.2 to 0.39 indicated fair agreement, and <0.2 indicated slight or very poor agreement.[9]

The κ value was calculated for endoscopists in each group based on the VS pattern and histology prediction. The diagnostic accuracy of prediction was calculated for each endoscopist according to the final histology. All analyses were performed using R version 2.15.1.

RESULTS

All seven participants completed both study stages. Baseline characteristics and endoscopic features of the 47 lesions are presented in Table 1. Of the 47 lesions, adenoma was the preoperative diagnosis for 35 lesions and carcinoma for 12, based on white light endoscopy. Lesion size was significantly larger in carcinomas than in adenomas ($p=0.001$). Ulcer lesion ($p=0.002$) and red color ($p=0.001$) were significant predictive factors for carcinoma, but depressed lesions did not have significant predictive value (Table 1).

Intraobserver data

The κ value for individual intraobserver variability varied

from 0.51 to 0.90 (Table 2). The mean κ value for intraobserver variability based on the Yao VS classification was 0.69 (experts, 0.74; trainees, 0.64) (Table 3). The mean κ values for both the expert and trainee groups indicated substantial agreement.

Interobserver data

The mean κ value for interobserver agreement based on VS classification was 0.45 (experts, 0.59; trainees, 0.43) in the first review and 0.40 (experts, 0.39; trainees, 0.38) in the second review. The mean κ value for interobserver agreement was 0.42 (experts, 0.49; trainees, 0.40) (Table 4). The mean κ values for both the expert and trainee group indicated moderate agreement.

Diagnostic accuracy of the VS classification

The diagnostic accuracy of the VS classification was calculated for each investigator according to the final histology. However, diagnostic accuracy was difficult to assess because of the retrospective analysis of the selected static images. The diagnostic accuracy of VS classification was 72.2% (experts, 69.2%; trainees, 75.3%) in the first review and 70.3% (experts, 69.2%; trainees, 71.5%) in the second review (Table 5). The mean diagnostic accuracy of the VS classification was 71.3% (experts, 69.2%; trainees, 73.4%).

DISCUSSION

An accurate diagnosis of gastric lesions is essential to determine optimal treatment. However, endoscopy with C-WLI

Table 1. Analysis of Baseline Characteristics and Endoscopic Features in Adenomas and Carcinomas

Characteristic	Adenoma ($n=35$)	Carcinoma ($n=12$)	p-value
Sex, male/female (% male)	28/7 (80.0)	9/3 (75.0)	0.700[a]
Age, yr	60.8±9.0	66.2±8.8	0.077
Diameter of lesion, mm	13.6±4.1	18.8±5.5	0.001
Lesion type			
Ulcer, with/without (% with)	2/33 (5.7)	6/6 (50.0)	0.002[a]
Type, depressed/not depressed (% depressed)	2/33 (5.7)	2/10 (16.7)	0.266[a]
Color, red/not red (% red)	9/26 (25.7)	10/2 (83.3)	0.001[a]

[a]Fisher exact test.

Table 2. The Intraobserver Agreement for Magnifying Endoscopy with Narrow Band Imaging Using Yao Vessel Plus Surface Classification

	E1	E2	T1	T2	T3	T4	T5
MV or MS	0.851	0.625	0.902	0.683	0.592	0.510	0.529
MV	0.628	0.719	0.415	0.619	0.551	0.384	0.465
MS	0.558	0.547	0.705	0.723	0.325	0.629	0.563
MV and MS	0.553	0.642	0.393	0.554	0.357	0.366	0.384

E, expert; T, trainee; MV, microvascular; MS, microsurface.

Table 3. The Intraobserver Agreement for Magnifying Endoscopy with Narrow Band Imaging Using Yao Vessel Plus Surface Classification

	E	T	Total
MV or MS	0.738	0.643	0.691
MV	0.674	0.487	0.580
MS	0.553	0.589	0.571
MV and MS	0.598	0.411	0.504

E, expert; T, trainee; MV, microvascular; MS, microsurface.

Table 4. The Interobserver Agreement for Magnifying Endoscopy with Narrow Band Imaging Using Yao Vessel Plus Surface Classification

	1st			2nd			Total (mean)		
	E1–E2[a]	T1–T5[b]	Total[b]	E1–E2[a]	T1–T5[b]	Total[b]	E1–E2[a]	T1–T5[b]	Total[b]
MV or MS	0.592	0.429	0.448	0.391	0.379	0.399	0.492	0.404	0.424
MV	0.576	0.365	0.378	0.506	0.292	0.347	0.541	0.329	0.363
MS	0.350	0.386	0.375	0.359	0.390	0.391	0.355	0.388	0.383
MV and MS	0.286	0.272	0.260	0.259	0.231	0.249	0.273	0.252	0.255

E, expert; T, trainee; MV, microvascular; MS, microsurface.
[a]Cohen κ; [b]Fleiss κ.

Table 5. The Diagnostic Accuracy of Magnifying Endoscopy with Narrow Band Imaging Using Yao Vessel Plus Surface Classification

	1st			2nd			Total (mean)		
	E	T	Total	E	T	Total	E	T	Total
Sensitivity	100.0	86.7	93.3	93.4	82.7	88.0	96.7	84.7	90.7
Specificity	54.7	70.0	62.4	57.8	66.3	62.0	56.3	68.1	62.2
Accuracy	69.2	75.3	72.2	69.2	71.5	70.3	69.2	73.4	71.3
PPV	51.0	59.8	55.4	53.0	57.5	55.3	52.0	58.7	55.3

E, expert; T, trainee; PPV, positive predictive value.

and pretreatment forceps biopsy is insufficient to obtain an accurate diagnosis.[1,10] ME-NBI is a powerful tool for characterizing the gastric mucosal surface because it enables the visualization of the precise microanatomies of both the MV and MS patterns of gastric mucosal lesions.[8]

Yao et al.[11] reported that the negative predictive value of the demarcation line was 100% and that an irregular MV pattern gave a diagnostic accuracy of 98.7% for small, flat gastric cancers in blinded prospective studies. Several studies have shown the usefulness of microscopic capillaries seen by ME-NBI for predicting gastric neoplasia among superficial depressed, flat, or elevated early gastric neoplastic lesions.[4-6]

Endoscopic diagnosis is subjective and therefore often shows relatively low inter-observer concordance. Our search of the literature revealed few studies of the intraobserver and interobserver variability of ME-NBI in diagnosing superficial gastric lesions. In this study, we performed a retrospective clinical investigation of the interobserver and intraobserver variability of experts and trainees analyzing ME-NBI images. A κ value >0.4 was considered acceptable.[9]

A previous study of ME-NBI images using clear, static images[12] is similar to our study. However, the criteria used to divide endoscopists into experts and trainees differed, and the proportion of depressed lesions was lower than that in our study. The lacking point is that which confirmed only agreement without comparison of C-WLI and ME-NBI.

In this study, the mean κ for the intraobserver agreement was 0.69 (substantial agreement) and the mean κ for interobserver agreement was 0.42 (moderate agreement). The Los Angeles classification is widely used for endoscopic gastroesophageal reflux disease and this classification has very low interobserver variability.[13,14] In comparison, the results from this study indicated a relatively reliable agreement.

In this retrospective study, the mean diagnostic accuracy of the VS classification was 71.3%. Considering that we made no distinction between depressed or nondepressed lesions and had a high proportion of nondepressed lesions, this result is acceptable. Using the Yao et al.[7] VS classification, we obtained reliable results with only brief instruction on image analysis. Interobserver agreement was not significantly different for experts and trainees.

This study had several limitations. Table 1 shows that C-WLI (diameter, mucosal colors, and ulcer) was highly accurate at predicting histology results. Thus, ME-NBI might not be necessary. However, this retrospective study evaluated observer agreement by random selection; therefore, comparing the accuracy of C-WLI and ME-NBI is not relevant. A multicenter study with many experts found that ME-NBI had a higher accuracy and specificity than C-WLI.[15]

Our results compared the mean κ values for interobserver agreement (Table 4) and found that the mean κ value slightly decreased in the second test, even without retraining. Agreement is considered to generally increase with training. The reduced agreement might be because ME-NBI is uncommon and endoscopists have little training using this method. Increased ME-NBI use and education might increase agreement. Contrary to expectations, the trainees had higher mean diagnostic accuracy than experts according to Table 5. Additional analysis may be necessary. Agreement for determination of a positive or negative diagnosis based on VS classification was higher than agreement for determinations made based on MV and MS patterns. This finding suggested that simultaneously

using both MS and MV patterns might increase agreement, although the possibility that higher agreement might have been due to simplified positive and negative categories cannot be ruled out.

Finally, there are few experienced endoscopists, and intraobserver and interobserver agreements were not satisfactory even in the expert group. We considered this result to be due to the fact that even if they are experts, there are many differences in the frequency of ME-NBI use.

This study was a single center retrospective study with several limitations. Future multicenter prospective and well-designed studies are needed.

In conclusion, by using the Yao VS classification, we obtained reliable results from experts and trainees after brief instruction. The classification provided an objective measure of ME-NBI for trainees who were not familiar with ME-NBI.

Conflicts of Interest

The authors have no financial conflicts of interest.

REFERENCES

1. Miwa K, Doyama H, Ito R, et al. Can magnifying endoscopy with narrow band imaging be useful for low grade adenomas in preoperative biopsy specimens? Gastric Cancer 2012;15:170-178.
2. Yao K. How is the VS (vessel plus surface) classification system applicable to magnifying narrow-band imaging examinations of gastric neoplasias initially diagnosed as low-grade adenomas? Gastric Cancer 2012;15:118-120.
3. Gono K, Obi T, Yamaguchi M, et al. Appearance of enhanced tissue features in narrow-band endoscopic imaging. J Biomed Opt 2004;9: 568-577.
4. Nakayoshi T, Tajiri H, Matsuda K, Kaise M, Ikegami M, Sasaki H. Magnifying endoscopy combined with narrow band imaging system for early gastric cancer: correlation of vascular pattern with histopathology (including video). Endoscopy 2004;36:1080-1084.
5. Yao K, Iwashita A, Tanabe H, et al. White opaque substance within superficial elevated gastric neoplasia as visualized by magnification endoscopy with narrow-band imaging: a new optical sign for differentiating between adenoma and carcinoma. Gastrointest Endosc 2008;68: 574-580.
6. Tsuji Y, Ohata K, Sekiguchi M, et al. Magnifying endoscopy with narrow-band imaging helps determine the management of gastric adenomas. Gastric Cancer 2012;15:414-418.
7. Yao K, Anagnostopoulos GK, Ragunath K. Magnifying endoscopy for diagnosing and delineating early gastric cancer. Endoscopy 2009;41: 462-467.
8. Schlemper RJ, Riddell RH, Kato Y, et al. The Vienna classification of gastrointestinal epithelial neoplasia. Gut 2000;47:251-255.
9. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med 2005;37:360-363.
10. Kim YJ, Park JC, Kim JH, et al. Histologic diagnosis based on forceps biopsy is not adequate for determining endoscopic treatment of gastric adenomatous lesions. Endoscopy 2010;42:620-626.
11. Yao K, Iwashita A, Tanabe H, et al. Novel zoom endoscopy technique for diagnosis of small flat gastric cancer: a prospective, blind study. Clin Gastroenterol Hepatol 2007;5:869-878.
12. Kaise M, Kato M, Urashima M, et al. Magnifying endoscopy combined with narrow-band imaging for differential diagnosis of superficial depressed gastric lesions. Endoscopy 2009;41:310-315.
13. Miwa H, Yokoyama T, Hori K, et al. Interobserver agreement in endoscopic evaluation of reflux esophagitis using a modified Los Angeles classification incorporating grades N and M: a validation study in a cohort of Japanese endoscopists. Dis Esophagus 2008;21:355-363.
14. Nasseri-Moghaddam S, Razjouyan H, Nouraei M, et al. Inter- and intra-observer variability of the Los Angeles classification: a reassessment. Arch Iran Med 2007;10:48-53.
15. Ezoe Y, Muto M, Uedo N, et al. Magnifying narrowband imaging is more accurate than conventional white-light imaging in diagnosis of gastric mucosal cancer. Gastroenterology 2011;141:2017-2025.