

# Reliability of Paris Classification for superficial neoplastic gastric lesions improves with training and narrow band imaging



## Authors

Helena Ribeiro<sup>1</sup>, Diogo Libânio<sup>2</sup>, Rui Castro<sup>2</sup>, Anibal Ferreira<sup>3</sup>, Pedro Barreiro<sup>4</sup>, Pedro Boal Carvalho<sup>5</sup>, Tiago Capela<sup>4</sup>, Pedro Pimentel-Nunes<sup>2,6,7</sup>, Cristina Santos<sup>7</sup>, Mário Dinis-Ribeiro<sup>2,7</sup>

## Institutions

- 1 Gastroenterology Department, Amato Lusitano Hospital, Castelo Branco, Portugal
- 2 Gastroenterology Department, Portuguese Oncology Institute of Porto, Porto, Portugal
- 3 Gastroenterology Department, Hospital of Braga, Braga, Portugal
- 4 Gastroenterology Department, Centro Hospitalar de Lisboa Ocidental, Lisboa, Portugal
- 5 Gastroenterology Department, Senhora da Oliveira Hospital, Guimarães, Portugal
- 6 Department of Surgery and Physiology, Porto Faculty of Medicine, Porto, Portugal
- 7 MEDCIDS/Porto Faculty of Medicine, Porto, Portugal

submitted 6.2.2018

accepted after revision 25.5.2018

## Bibliography

DOI <https://doi.org/10.1055/a-0828-7541> |  
Endoscopy International Open 2019; 07: E633–E640  
© Georg Thieme Verlag KG Stuttgart · New York  
eISSN 2196-9736

## Corresponding author

Helena Sofia Brito Ribeiro, Gastroenterology Department, Hospital Amato Lusitano, Avenida Pedro Álvares Cabral, 6000-085 Castelo Branco, Portugal  
Fax: +351 272 000 272  
[helena.brito@brito@gmail.com](mailto:helena.brito@brito@gmail.com)



Appendix 1  
Online content viewable at:

<https://doi.org/10.1055/a-0828-7541>

## ABSTRACT

**Background and study aims** Paris Classification is used to classify gastrointestinal superficial neoplastic lesions and to predict presence of submucosal invasion. We aimed to evaluate interobserver reliability and agreement for this classification among Western endoscopists.

**Methods** A total of 54 superficial gastric lesions were independently classified according to Paris classification by eight endoscopists (4 experts and 4 non-experts). Observers were asked to classify two sets of images – first, obtained with high-resolution white light (HR-WL) endoscopy and secondly, with the same HR-WL images paired with images obtained with high-resolution Narrow Band Imaging (HR-NBI) – HR-WL + NBI image group.

**Results** Overall interobserver reliability when asked to classify in I, II or III was good both using HR-WL images and HR-WL + NBI images (wK of 0.65 and 0.70, respectively). The proportion of agreement for type III lesions was 0.48 for HR-WL images increasing to 0.74 in the HR-WL + NBI group. Interobserver reliability for identification of a IIc component was only moderate (wK 0.47). NBI improves both sensitivity and interobserver reliability among trainees (from wK 0.19 to 0.47). Specificity was higher than sensitivity in predicting submucosal invasion.

**Conclusion** Overall, the reliability of Paris classification is moderate to good. Training on this classification or its revision and use of technology such as NBI may improve not only reliability and agreement but also accuracy.

## Introduction

Gastric superficial neoplastic lesions are defined as lesions with an endoscopic appearance suggestive of invasion limited to the mucosa or submucosa [1–2]. They include low- or high-grade noninvasive neoplasia and adenocarcinoma with no evidence of deep submucosal invasion [3]. Recognition and detection of

these early lesions is vital to improve survival of patients with gastric cancer, which is the fifth most common malignancy worldwide and the third leading cause of cancer death [4].

The Paris Classification is an international standard for endoscopic classification of gastrointestinal superficial neoplastic lesions, adapted from the Japanese macroscopic classification for gastric cancer [1, 5]. Japanese studies demonstrated that the

different types and subtypes of the Paris classification are predictive factors of the extent of invasion into the submucosa, which correlates with risk of nodal metastases in gastric lesions [1]. Indeed, Paris 0-I, 0-IIc and 0-III are associated with a higher risk of submucosal invasion (57%, 37% and 40%, respectively) when compared with 0-IIa and 0-IIb lesions (29% and 20%, respectively) [1]. Therefore, the Paris Classification became an important factor to be considered in endoscopic assessment of superficial lesions as it helps to predict feasibility and curability of endoscopic resection and also to choose the more adequate endoscopic resection technique, along with other features, such as lesion size [1, 6–7]. Superficial gastric lesions should be described in accord with the Paris classification after endoscopic evaluation with high-resolution white-light (HR-WL) endoscopy and with high-resolution narrow-band imaging (HR-NBI) [8]. HR-NBI is highly accurate for diagnosis of early gastric neoplasia [9]. It improves characterization of mucosal surface and margins of gastrointestinal lesions, and so, it may play a role in assisting endoscopists in classifying gastric lesions according to the Paris classification [10].

Despite the important role of the Paris Classification in management of patients with superficial gastric neoplasia and its complexity, data on the reproducibility of this classification among endoscopists are scarce. Thus, we performed a multicenter study to evaluate interobserver reliability and agreement for the Paris Classification of superficial neoplastic gastric lesions among Western endoscopists with different levels of expertise and the influence of HR-NBI in this reliability.

## Methods

### Gastric lesion selection

Images of gastric lesions were collected from a pool of consecutive endoscopic images from gastric endoscopic submucosal dissections (ESD) performed between January 2015 and April 2017 at the Portuguese Oncology Institute of Porto. Endoscopic procedures were performed with Olympus GIF-HQ190 endoscopes (with dual-focus) and EVIS EXERA III video processor. Lesions selected were those with a paired HR-WL endoscopic image and corresponding HR-NBI endoscopic image. Superficial lesions ineligible for endoscopic resection were also considered when HR-WL and NBI images were available, to fulfil the overall spectrum of type-0 gastric lesions. A sample of 54 lesions was obtained. Two of the 54 selected lesions were not eligible for endoscopic resection and were submitted to surgical resection. In all other cases, ESD was performed as first treatment. Each lesion had two endoscopic images: one HR-WL image plus one corresponding HR-NBI image.

### Selection of endoscopists

A group of eight Portuguese endoscopists was selected to classify the selected images. The endoscopists had different levels of expertise in gastric ESD: four experts (>100 exams performed; MDR; PPN; PB; AF), two beginners (<20 ESD, under supervision; PBC; TC) and two trainees (see experts at work; DL; RC). Two experts and the two trainees worked in the same hospital (MDR; PPN; DL; RC) and the other endoscopists were from

four different hospitals. The trainees and the beginners first trained (in ESD and in the Paris classification application) with two of the experts (MDR and PPN).

### Lesion classification process

Two online forms (► Fig. 1a, ► Fig. 1b) containing the 54 lesions were sent to the endoscopists at two different times. The first form consisted of one image obtained with HR-WL endoscopy for each lesion: the –HR-WL image group. The endoscopists had to classify each lesion as type 0-I, 0-II or 0-III and then specify the subtype. They were also asked if there were features predictive of submucosal invasion and to estimate the lesion diameter in millimeters. Predictive features of deep submucosal invasion included marked depression, markedly elevated margins, interruption of gastric folds, and absence of mucosal pattern. Endoscopists were asked to state if there were any features of submucosal invasion (yes/no). Two weeks after, a second form was sent with the same 54 lesions paired with a corresponding NBI image (2 images per lesion, 1 HR-WL image plus a HR-NBI image), in a different order – HR-WL + NBI image group. The endoscopists were again asked the same questions as previously described.

### Statistical analysis

Considering categorical variables, interobserver agreement among endoscopists was assessed using the proportions of agreement (PA) and proportion of specific agreement (specific PA for each category), as recommended by the “Guidelines for reporting reliability and agreement studies (GRRAS)” [11]. A PA equal to 0.5 means that when an observer attributes a certain classification, there is a 50% probability that another observer will attribute the same classification. If the lower limit of the 95% confidence interval (CI) for PA was under 0.50, agreement was considered poor [12]. The proportions of agreement relative to each individual category (proportion of specific agreement) help to understand that agreement is high in some categories and low in others. Specific PA for category A estimates the conditional probability, given that one of the raters, randomly selected, makes a rating on category A, the other rater will also do so. Reliability was evaluated with the weighted kappa ( $wkappa$ ) or kappa statistic ( $k$ -Light’s kappa for  $n$  raters). Kappa adjusts PA to the agreement expected by chance, so the distribution of ratings in the different classes influences the results. Consequently, it is possible to obtain a high proportion of agreement and a low kappa when prevalence of a given rating is very high or very low [13]. Kappa values below 0.20 were considered as slight reliability; those ranging between 0.21 and 0.40 as fair reliability, those between 0.41 and 0.60 as moderate reliability, those between 0.61 and 0.80 as substantial reliability, and values larger than 0.80 as almost perfect reliability [14]. Considering continuous variables, reliability was assessed with Intraclass Correlation Coefficient (ICC) and interobserver agreement among endoscopists was assessed using the Information Based Measure of Disagreement (IBMD) [15]. ICC ranges from 0 (no reliability) to 1 (perfect reliability), on the other hand IBMD ranges from 0 (no disagreement or perfect agreement) to 1 (perfect disagreement or no agreement).

## Paris Classification of Superficial Gastric Lesions

• Required

Figure 12

According to the Paris Classification, please classify the following lesion as type I, type II or type III. \*



- I
- II
- III

Please classify now the gastric lesion according to the Paris Classification. \*

- Is
- Ip
- Isp
- Ila
- Iib
- Iic
- Iic+Ila
- Ila+Iic
- Ila+Iib
- Iib+Ila
- III+Iic
- Iic+III
- III
- Other:

Indicate your level of confidence in your previous classification. \*

- High
- Low

In this lesion is there any feature predictive of submucosal invasion? \*

- Yes
- No

Indicate the estimate diameter of the lesion (in millimeters) \*

Your answer \_\_\_\_\_

BACK

NEXT

Page 13 of 55

Never submit passwords through Google Forms.

## Paris Classification of Superficial Gastric Lesions - part 2 (HD+NBI)

• Required

Figure 36

According to the Paris Classification, please classify the following lesion as type I, type II or type III. \*



- I
- II
- III

Please classify now the gastric lesion according to the Paris Classification. \*

- Is
- Ip
- Isp
- Ila
- Iib
- Iic
- Iic+Ila
- Ila+Iic
- Ila+Iib
- Iib+Ila
- III+Iic
- Iic+III
- III
- Other:

Indicate your level of confidence in your previous classification. \*

- High
- Low

In this lesion is there any feature predictive of submucosal invasion? \*

- Yes
- No

Indicate the estimate diameter of the lesion (in millimeters) \*

Your answer \_\_\_\_\_

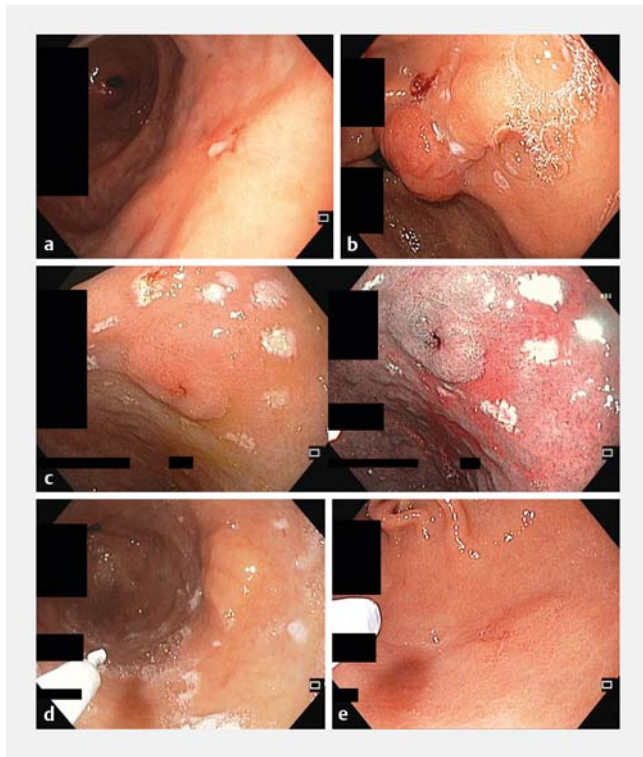
BACK

NEXT

Page 37 of 55

Never submit passwords through Google Forms.

► Fig. 1 Example of image evaluation with online forms using “Google form”. a HR-WL image of a superficial gastric lesion. b HR-WL + HR-BI image of a superficial gastric lesion.



► **Fig. 2** Example of images of the online forms and respective classifications attributed by the observers. **a** HR-WL with classification of a type II lesion by observers PB, AF, PBC, RC and TC versus type III lesion by observers MDR, DL, and PPN. **b** HR-WL with classification as a subtype Is lesion by observers MDR, PPN, PB, AF, PBC and RC versus type IIa by observers DL and TC. **c** HR-WL + HR-NBI with classification as a subtype IIa by observers MDR, DL, PPN, AF and PBC versus IIa + IIc by observers PB, RC and TC. **d** HR-WL with classification as a type II and subtype IIa lesion by all the observers. **e** HR-WL with classification as a type II and subtype IIb lesion by all the observers.

The terms “reliability” and “agreement” are conceptually distinct terms. Reliability can be defined as the ability of a measurement to differentiate between subjects. On the other hand, agreement is the degree to which scores or ratings are identical [11].

Both concepts are important, as they provide information about the quality of measurements. R software was used to compute the PA, Wkappa, kappa, ICC and IBMD with “obs. Agree” and “psy” packages, respectively. Ninety-five percent CIs (95%CI) were calculated for all measures.

## Results

### Type 0-I, -II and III lesions and subtypes

Total interobserver agreement and reliability for the Paris classification among the 8 endoscopists for the categories type 0-I, -II or -III lesions was good in the HR-WL image group, with a weighted kappa (wK) of 0.65 and a proportion of agreement (PA) of 91%. Results were similar in the HR-WL + NBI image group (wK 0.70; PA 93%). ► **Fig. 2** shows an example of images of the online forms and the respective classifications attributed

by the observers are shown. Appendix 1 shows the classification attributed by each observer to each of the superficial gastric lesions and the pathological depth of the 54 lesions.

Considering each category individually, in the HR-WL image group, the PA between endoscopists was 0.75 for type 0-I, 0.95 for the type 0-II and 0.48 for the type III lesions. In the HR-WL + NBI image group the PA for the type 0-I and II were similar to those of the HR-WL images group (0.70 and 0.96, respectively). In contrast, the PA for type III lesions increased with HR-WL + NBI when compared with HR-WL (PA 0.74 vs 0.48), however, without statistical significance. Regarding levels of expertise, total interobserver reliability for categories type 0-I, -II or -III lesions for both images groups was good among the experts and among the beginners (wK 0.72 and 0.77, respectively). Among the trainees, total interobserver reliability was fair (wK 0.33) in the HR-WL image group and increased to moderate (wK 0.60) in the HR-WL + NBI image group. The trainees agreed less about the type III lesions compared with types II and I, however, without statistical significance (► **Table 1**).

Overall interobserver reliability among all endoscopists in classification of the subtype IIc lesions was moderate and did not improve significantly with the addition of the HR-NBI images (wK 0.47 and wK 0.50 respectively). On the other hand, considering just the trainees endoscopists, there was poor interobserver reliability with HR-WL that increased with HR-WL + NBI (from 0.19 to 0.47), however, without statistical significance (► **Table 2**).

### Lesion size

Regarding estimation of lesion size, both beginners and trainees had significantly more disagreement among them compared with the expert endoscopists (IBMD of 0.322 [0.275,0.374], 0.320 [0.276,0.369] and 0.236 [0.214, 0.262], respectively) in the HR-WL image group. In the HR-WL + NBI image group the IBMD decreased in both the beginner and trainee groups (IBMD of 0.243 [0.198,0.291; 0.276 [0.230,0.323], respectively), and beginners, trainees, and experts did not differ significantly considering the disagreement among them.

► **Fig. 3** shows the diameter estimation for each lesion made by the endoscopists.

### Submucosal invasion

Considering the histology analysis, 1.9% of the lesions were sm1 and 13% were sm2 lesions. Overall reliability among the eight endoscopists for existence of endoscopic features predicting submucosal invasion was moderate and the beginners had the lowest overall agreement – fair in both image groups. Considering histology as the gold standard for submucosal invasion, the observers had higher specificity than sensitivity in predicting submucosal invasion (ranging from 96% for the beginners and the trainee groups and 83% for the group of experts from different institutions). The beginners and trainees had the lowest sensitivity in predicting submucosal invasion in the HR-WL images group (sensitivity of 38%) but with the addition of the HR-NBI, sensitivity increased to 50% in the beginner group and to 63% in the trainee group. In contrast, with NBI,

► **Table 1** Interobserver agreement and agreement for the Paris classification among the endoscopists for categories type 0-I, -II or -III lesions.

	HR-WL image group		Wkappa [95%CI]	HR-WL + NBI image group
	Wkappa [95%CI]	PA [95%CI]		
All endoscopists (n = 8 observers)				
Lesion:	0.65 [0.45,0.82]	0.91 [0.87,0.95]	0.70 [0.48,0.88]	0.93 [0.88,0.97]
I		0.75 [0.37,0.89]		0.70 [0.28,0.87]
II		0.95 [0.92,0.97]		0.96 [0.93,0.98]
III		0.48 [0.09,0.74]		0.74 [0.00,1.00]
Experts endoscopists (n = 4 observers)				
Lesion:	0.72 [0.34,0.89]	0.92 [0.87,0.97]	0.72 [0.43,0.91]	0.94 [0.87,0.98]
I		0.84 [0.40,1.00]		0.70 [0.22,0.91]
II		0.95 [0.92,0.98]		0.96 [0.93,0.99]
III		0.47 [0.00,0.86]		0.78 [0.33,1.00]
Beginners endoscopists (n = 2 observers)				
Lesion:	0.77 [0.34,1.00]	0.94 [0.87,1.00]	0.66 [0.27,0.94]	0.92 [0.85,0.98]
I		0.75 [0.00,1.00]		0.67 [0.00,1.00]
II		0.97 [0.92,1.00]		0.96 [0.91,0.99]
III		0.80 [0.00,1.00]		0.67 [0.00,1.00]
Trainees endoscopists (n = 2 observers)				
Lesion:	0.33 [0.00,0.66]	0.85 [0.76,0.94]	0.60 [0.24,0.88]	0.89 [0.80,0.96]
I		0.50 [0.00,0.89]		0.61 [0.22,0.89]
II		0.92 [0.86,0.98]		0.93 [0.88,0.98]
III		0.00 [0.00,0.00]		0.67 [0.00,1.00]

the experts had lower sensitivity (37%) compared with the sensitivity obtained in the HR-WL image group (sensitivity of 85%).

## Discussion

In this study, we showed for the first time the reliability of Paris Classification among Western endoscopists with different expertise and with NBI. The results were reasonable and better between experts than between inexperienced observers and showed improvement with NBI.

The classification includes 3 categories – protruding lesions (type 0-I), nonprotruding and nonexcavated lesions (type 0-II) and excavated lesions (type 0-III). Each of these categories have subtypes and may also be considered mixed patterns. The most frequent gastric superficial lesions are type 0-IIc component, whereas the type 0-IIb and type 0-III are rare [1, 5].

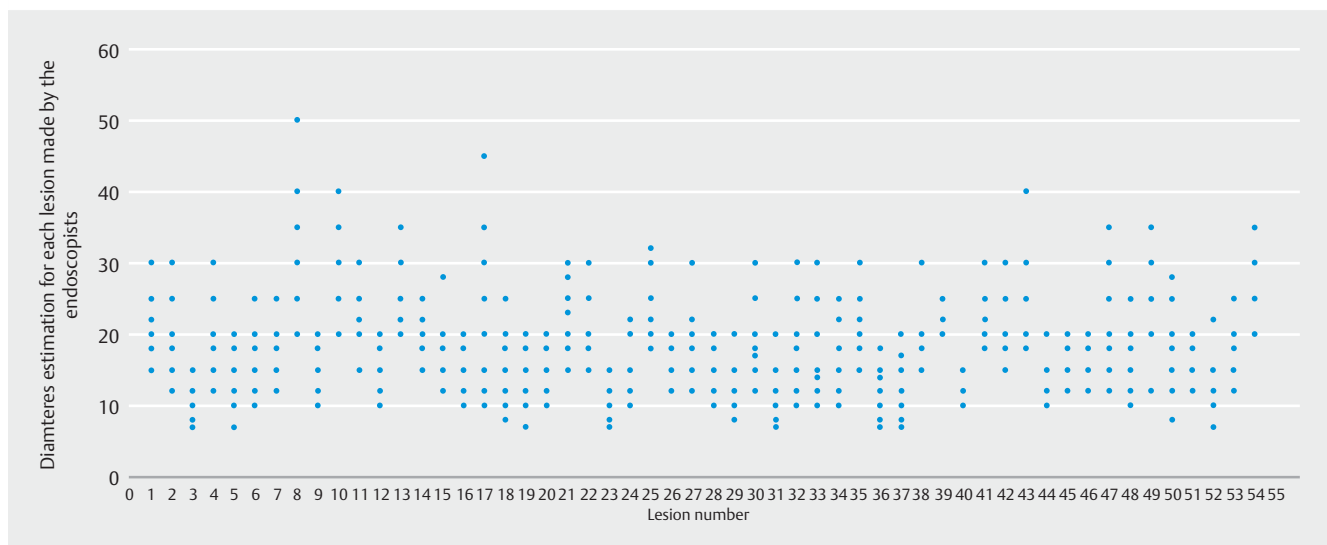
In our study, the participants presented good overall interobserver agreement and reliability in classifying the gastric lesions as type 0-I, II or III, with or without NBI images. Morphology of a type 0 lesion has predictive value for submucosal invasion and for associated risk of lymph node (LN) metastases. According to surgical Japanese series, risk of invasion into the submucosa is higher in type 0-I or depressed 0-II c lesions [1]. A

more recent meta-analysis demonstrated that lesions that were macroscopically depressed (type 0-IIc lesions, type 0-III lesions, and lesions with one of these components) were related to LN metastasis in gastric cancer limited to the mucosa [16]. Although Paris classification is frequently used to select the more adequate endoscopic resection technique and to predict probability of submucosal invasion, the evidence concerning interobserver variability for this classification in gastric superficial lesions is scarce. In our study, the PA classifying type I lesions was the highest (0.95). However, observer agreement in classifying the depressed lesions was not so favorable. The PA for type III was the lowest (0.48) with HR-WL images and overall interobserver agreement when they were asked to identify a IIc component was only moderate (0.47). These facts may impair the clinical relevance of the classification in identifying lesions with higher risk of submucosal invasion and LN metastasis.

The HR-NBI image may play an important role in this matter. In fact, with HR-NBI images, the PA for type III lesions increased from 0.48 to 0.74 among all the endoscopists and interobserver reliability in classification of the subtype IIc also increased considerable for trainees (from 0.19 to 0.47) with HR-NBI. HR-NBI also improved reliability among the trainees from fair (0.33) to moderate (0.60) when they were asked to classify

► **Table 2** Interobserver agreement and reliability among all endoscopists for classification of the subtype IIc.

	HR-WL image group		HR-WL + NBI image group	
	wkappa	PA	wkappa	PA
All endoscopists (n = 8 observers)				
Gastric lesion:	0.47 [0.36,0.61]	0.74 [0.67,0.79]	0.50 [0.40,0.62]	0.75 [0.69,0.80]
IIC		0.77 [0.67,0.83]		0.77 [0.69,0.83]
Not IIC		0.70 [0.61,0.78]		0.73 [0.63,0.80]
Experts endoscopists (n = 4 observers)				
Gastric lesion:	0.56 [0.40,0.70]	0.78 [0.70,0.84]	0.56 [0.43,0.71]	0.78 [0.70,0.85]
IIC		0.78 [0.66,0.85]		0.79 [0.70,0.86]
Not IIC		0.77 [0.68,0.85]		0.76 [0.65,0.85]
Beginners endoscopists (n = 2 observers)				
Gastric lesion:	0.43 [0.22,0.65]	0.72 [0.59,0.83]	0.46 [0.26,0.67]	0.72 [0.59,0.83]
IIC		0.78 [0.64,0.87]		0.76 [0.61,0.86]
Not IIC		0.63 [0.43,0.78]		0.67 [0.50,0.81]
Trainees endoscopists (n = 2 observers)				
Gastric lesion:	0.19 [-0.05,0.47]	0.61 [0.46,0.74]	0.47 [0.18,0.67]	0.74 [0.61,0.83]
IIC		0.68 [0.51,0.79]		0.77 [0.61,0.86]
Not IIC		0.51 [0.31,0.68]		0.71 [0.54,0.84]



► **Fig. 3** Diameter estimates made by the endoscopists for each lesion.

type I, II or III lesions. Among the experts and beginners, HR-NBI did not have this impact, perhaps because they already had good results with the HR-WL images. This study aimed to estimate the general reliability of the Paris classification among endoscopists and to discuss differences according to different technologies and level of training. To overcome the limited size of the sample, we took into account the prevalence of outcomes, spectrum of changes, and number of repetitions (a product of number of images and observers). Thus, we believe

it seems reasonable to conclude that the HR-NBI may be more helpful for less-experienced endoscopists and important in the learning process.

The relevance of macroscopic appearance of early gastric cancer (EGC) may also be useful for prediction of histological differentiation and clinical behavior, particularly in differentiated EGC [17]. Elevated lesions are more common in well and moderately differentiated cancer, type IIb is more common in signet-ring-cell carcinoma and type IIc and III in poorly differen-

tiated cancer [17]. ESD is the treatment of choice for most gastric superficial neoplastic lesions. Presence of ulceration in gastric lesions is a risk factor for non-curative endoscopic resection [6, 18]. Large lesion size is another important risk factor for LN metastasis in mucosal and submucosal EGG and for non-curative ESD [6]. When asked to assess lesion size, disagreement was higher among beginners and trainees in the HR-WL image group, but improved in the HR-WL + NBI group.

Other endoscopic factors have been reported as predictive of non-curative ESD, such as localization in the upper stomach, a non-nodular surface and presence of fusion gastric folds [6–7, 19]. When endoscopists were asked to say whether lesions had features suggesting submucosal invasion, they based their answers on other features besides Paris Classification and size, despite the fact that each feature suggestive of submucosal invasion was not classified independently. Overall reliability among the endoscopists for predicting submucosal invasion was moderate. They were better at detecting lesions that did not actually have submucosal invasion than lesions with truly submucosal invasion. HR-NBI increases sensitivity in predicting submucosal invasion among less-experienced endoscopists but decreases it in the experts.

Vindigni et al. assessed interobserver reliability and agreement on the Paris Classification of superficial gastric lesions between Italian and Japanese endoscopists. In that study, the authors verified that interobserver reliability was only moderate ( $\text{Kappa}=0.54$ ) [20]. Similar results were achieved in another study focused on assessment of colonic lesions, which also demonstrated only moderate interobserver agreement and reliability among international Western experts for the Paris classification system, with a kappa of 0.42 [21]. Those authors concluded that high interobserver variability renders use of this classification in clinical practice questionable. Their results were similar to ours as they also achieved moderate interobserver agreement and reliability.

A limitation of this study was the fact that the lesions were assessed based on still images instead of videos. Besides, there was only one HR-WL image and one image with HR-NBI for each lesion, with heterogeneous qualities and perspectives. These factors could have impaired the endoscopists' assessment and could have resulted in an underestimation of interobserver agreement between them. Nevertheless, this setting is also closer to the clinical real world. Also, it was not possible to assess features that could help predict submucosal invasion in every lesion, such as precise lesion location in the stomach, pliability and movement. We only asked the endoscopists to indicate whether the lesions had characteristics of submucosal invasion (yes or no), and not what characteristics led to their decision. However, there are no studies on this matter.

## Conclusion

In conclusion, interobserver agreement and reliability of Paris classification is moderate to good and is higher among experts. HR-NBI seems to improve reliability and agreement among less experienced endoscopists, but further studies with larger sam-

ples are needed to ascertain the value of the images and compare their performance with conventional chromoendoscopy.

## Competing interests

None

## References

- [1] The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002. *Gastrointest Endosc* 2003; 58: S3–S43
- [2] Japanese Gastric Cancer Association. Japanese classification of gastric carcinoma: 3rd English edition. *Gastric Cancer* 2011; 14: 101–112
- [3] Pimentel-Nunes P, Dinis-Ribeiro M, Ponchon T et al. Endoscopic submucosal dissection: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. *Endoscopy* 2015; 47: 829–854
- [4] Ferlay J, Soerjomataram I, Dikshit R et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015; 136: E359–E386
- [5] Endoscopic Classification Review Group. Update on the Paris Classification of Superficial Neoplastic Lesions in the Digestive Tract. *Endoscopy* 2005; 37: 570–578
- [6] Kim EH, Park JC, Song IJ et al. Prediction model for non-curative resection of endoscopic submucosal dissection in patients with early gastric cancer. *Gastrointest Endosc* 2017; 85: 976–983
- [7] Toyokawa T, Inaba T, Omote S et al. Risk factors for non-curative resection of early gastric neoplasms with endoscopic submucosal dissection: Analysis of 1,123 lesions. *Exp Ther Med* 2015; 9: 1209–1214
- [8] Pimentel-Nunes P, Libânio D, Dinis-Ribeiro M. Evaluation and management of gastric superficial neoplastic lesions. *GE - Port J Gastroenterol* 2017; 24: 8–21
- [9] Pimentel-Nunes P, Dinis-Ribeiro M, Soares J et al. A multicenter validation of an endoscopic classification with narrow band imaging for gastric precancerous and cancerous lesions. *Endoscopy* 2012; 44: 236–246
- [10] Boeriu A, Boeriu C, Drasovean S et al. Narrow-band imaging with magnifying endoscopy for the evaluation of gastrointestinal lesions. *World J Gastrointest Endosc* 2015; 16: 110–120
- [11] Kottner J, Audigé L, Brorson S et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011; 64: 96–106
- [12] Grant JM. The fetal heart rate trace is normal, isn't it? Observer agreement of categorical assessments *Lancet* 1991; 337: 215–218
- [13] Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990; 43: 543–549
- [14] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174
- [15] Henriques T, Antunes L, Bernardes J et al. Information-based measure of disagreement for more than two observers: a useful tool to compare the degree of observer disagreement. *BMC Med Res Methodol* 2013; 13: 47
- [16] Kwee RM, Kwee TC. Predicting lymph node status in early gastric cancer. *Gastric Cancer* 2008; 11: 134–148
- [17] Jung DH, Park YM, Kim JH et al. Clinical implication of endoscopic gross appearance in early gastric cancer: Revisited. *Surg Endosc Other Interv Tech* 2013; 27: 3690–3695
- [18] Ferlay J, Shin HR, Bray F et al. Estimates of worldwide burden of cancer in 2008: GLOBOCAN. *Int J Cancer* 2010; 127: 2893–2917

- [19] Hirasawa K, Kokawa A, Oka H et al. Risk assessment chart for curability of early gastric cancer with endoscopic submucosal dissection. *Gastrointest Endosc* 2011; 74: 1268 – 1275
- [20] Vindigni C, Marini M, Cevenini G et al. Italy-Japan agreement and discrepancies in diagnosis of superficial gastric lesions. *Front Biosci (Elite Ed)* 2010; 2: 733 – 738
- [21] van Doorn SC, Hazewinkel Y, East JE et al. Polyp morphology: an interobserver evaluation for the Paris classification among international experts. *Am J Gastroenterol* 2015; 110: 180 – 187