

ABSTRACT

Background and aims.

Patients with longstanding ulcerative colitis (UC) are at increased risk of developing colorectal neoplasia. Chromoendoscopy (CE) increases detection of lesions, and Kudo pit pattern classification I and II have been suggested to be predictive of benign polyps in UC. Little is known on the use of this classification in non-magnified high definition (HD) (virtual) CE and Narrow Band Imaging (NBI), or on the inter-observer agreement. The aim of this pilot study was to assess the diagnostic accuracy and the inter-observer agreement of the Kudo pit pattern classification in UC patients undergoing surveillance with methylene blue CE or NBI in a multicenter study.

Methods.

Fifty images of lesions identified in 27 UC patients (13 neoplastic) either with classical CE (methylene blue 0.1%) (n=24) or NBI (n=26), were selected by an independent investigator. Images were selected from a randomized controlled trial to compare CE and NBI. All non-magnified images were obtained with an Exera II Olympus processor and were mounted in a PowerPoint file in a standardized way (same size; black background). Ten endoscopists with extensive experience in NBI/CE were asked to assess the lesions for the predominant Kudo pit pattern (I, II, III_L, III_S, IV and V), to indicate if they thought the lesion was neoplastic and how confident they were about the diagnosis. Histology was used as the gold standard.

Results.

Median sensitivity, specificity, negative predictive value and positive predictive value for diagnosing neoplasia based on the presence of pit pattern other than I or II was 77%, 68%, 88% and 46%, respectively. Diagnostic accuracy was significantly higher when a diagnosis was made with a high level of confidence (77% vs 21%, $p < 0.001$). The overall inter-observer agreement for any pit pattern was only fair ($\kappa = 0.282$), with CE being significantly better than NBI (0.322 vs. 0.224, $p < 0.001$). From a clinical viewpoint, the difference between neoplastic and non-neoplastic lesions is important. The agreement for differentiation between non-neoplastic patterns (I, II) and neoplastic patterns (III_L, III_S, IV, or V) was moderate ($\kappa 0.587$) and even significantly better for NBI in comparison to CE ($\kappa 0.653$ vs. 0.495, $p < 0.001$).

Conclusions.

Differentiation between non-neoplastic and neoplastic pit patterns in UC lesions shows a moderate to substantial agreement among expert endoscopists. The agreement for differentiating neoplastic from non-neoplastic lesions is significantly better for NBI in comparison to HDCE. The assessment of pit pattern I or II with non-magnified HD-CE or NBI has a high negative predictive value to rule out neoplasia.

1 INTRODUCTION

2 Patients with longstanding ulcerative colitis (UC) are at increased risk of developing colorectal cancer
3 (CRC).^{1,2} Recent epidemiological data suggest that the risk of CRC in those patients is decreasing, but
4 still higher than an average risk population without inflammatory bowel disease.³ In comparison to
5 older data, this decrease in CRC is often attributed to both a better medical treatment and better
6 surveillance. Indeed, the risk of UC-associated CRC increases with a longer disease duration and greater
7 disease activity and extent.⁴ In addition, better endoscopic surveillance allows early detection of
8 dysplasia prior to CRC development and allows curative endoscopic or surgical resection.

9 Chromo-endoscopy (CE) increases the detection of neoplasia in patients with longstanding UC, in
10 comparison to regular white light endoscopy with random four quadrant biopsies taken every 10
11 centimetres.⁵⁻⁷ Because of the significantly increased diagnostic yield, CE is now the preferred
12 surveillance method in several guidelines.^{8,9} In a randomized controlled trial comparing methylene
13 blue CE and 4 quadrant random biopsies during white light endoscopy (WLE), a pragmatic Kudo pit
14 pattern classification was put forward that allowed to differentiate in a dichotomous way between
15 non-neoplastic and neoplastic lesions in patients with longstanding colitis undergoing surveillance.⁵ As
16 such, Kudo I and II have been suggested to be predictive of benign lesions. However, the diagnostic
17 accuracy and inter-observer variability of non-magnified high definition (HD) CE is unknown.

18 The role of Narrow Band Imaging (NBI) in surveillance of longstanding colitis is still controversial, mainly
19 because NBI has not been shown to detect more dysplasia in comparison to HD-CE or HD-WLE, and
20 NBI use is not supported by international consensus guidelines.¹⁰⁻¹³ Furthermore, it's unclear whether
21 the pit pattern classification can also be applied to HD NBI.

22 The aim of the current pilot study was to assess diagnostic accuracy and the inter- and intra-observer
23 agreement of Kudo Pit Pattern in UC surveillance endoscopy with non-magnified HD-CE or HD-NBI..

1 **METHODS**

2 **Selection of endoscopic images**

3 Stored images from endoscopic procedures performed between July 2008 and March 2012 were
4 retrieved from a randomized controlled trial comparing HD-CE versus NBI for the detection of
5 neoplasia in patients with longstanding UC (Clinicaltrials.gov NCT01882205).¹² Written informed
6 consent was obtained from all patients participating in this trial. The images were stored on a
7 computerized database of the Endoscopy Unit of the University Hospitals Leuven (Leuven, Belgium).
8 One investigator (MF), who was not involved in the endoscopic procedures nor in the assessment of
9 the pit patterns in a later phase of the study, screened all 267 available images, and selected 50 of
10 them to be used for the study purpose. The images were selected to ascertain the image was sharp
11 with a clearly visible lesions with visible pits and crypts. In addition, these images had previously been
12 scored for Kudo pit pattern by a single experienced gastroenterologist (RB) during the endoscopic
13 procedure (so blinded to the final pathology) and histological samples of matching biopsies had been
14 analysed by an experienced pathologist. The intend was to have an equal distribution between CE and
15 NBI pictures and a selection comprising a wide range of the Kudo pit pattern (I: 8/50; II: 22/50; III L
16 9/50; IIIs 5/50; IV:6/50) and morphology (see table 1). The original histological diagnosis was non-
17 neoplastic in 37 (2 mild architectural abnormalities, 4 inactive chronic colitis, 6 UC inflammatory
18 changes, 5 inflammatory pseudopolyps, 20 hyperplastic polyps), and neoplastic in 13 (historically
19 and per protocol defined as 6 adenoma with low grade dysplasia, 5 adenoma-like mass (ALM) with low
20 grade dysplasia, 1 dysplasia-associated lesion or mass (DALM) with low grade dysplasia, and 1 sessile
21 serrated adenoma with low grade dysplasia). Lesions were classified according to the Vienna
22 classification¹⁴. All biopsies were reviewed by an pathologist with specific expertise in gastrointestinal
23 pathology. In case of dysplasia, the diagnosis was confirmed by a second expert pathologist.

1 The selected images were originating from 27 different patients with longstanding UC undergoing
2 surveillance endoscopy [9/27 (33%) female, median (interquartile range, IQR) disease duration at index
3 endoscopy 16 (10-22) years]. The maintenance therapy consisted of mesalamine in 18 (67%), a
4 thiopurine in 7 (26%), and/or a biologic in 11 (41%) patients. Twenty-four images (44%) were retrieved
5 from 12 patients who underwent surveillance endoscopy with classical CE, using methylene blue 0.1%.
6 The other 26 images (56%) were derived from 15 patients undergoing surveillance colonoscopy with
7 NBI.

8 Images were obtained with an Exera II processor in high definition format, and high definition
9 colonoscopes (180Q series Olympus, Tokyo) and downloaded from the server in a joint photographic
10 experts group or JPEG format with a size of about 1000 kilobyte and a pixel array of 1008 x 1280 and
11 72 dots per inch. Next, all 50 images were mounted in a Power point file in two different sets with a
12 random order, but in a standardized way. All images received a black background in the PowerPoint
13 file, and were further displayed in their original size, without changing the brightness, contrast or
14 colour balance with any software. Figure 1 shows an example of two of these images.

15 **Assessment of endoscopic images**

16 Ten endoscopists (RB, TB, AP, PB, MR, KR, ED, JEE, AI, and ES) with previous experience in NBI and HD-
17 CE, as well as Kudo pit pattern classification, through participation in clinical trials were invited for this
18 study. They received twice a PowerPoint file with the same 50 images as well as a score sheet to
19 indicate the most advanced Kudo pit pattern, the neoplastic nature and how confident they were
20 about their assessment on a scale from 1 (not sure at all) to 5 (very sure) (See Supplementary File 1).
21 Score 4 and 5 were considered as a high-level confidence diagnosis. Pit pattern diagnosis was made
22 according to the Kudo classification as types I, II, III_L, III_S, IV, or V (See Supplementary File 2).¹⁵ The
23 interval between set 1 and 2 was at least 10 weeks to minimize recall bias. Furthermore, the

endoscopists (except for RB and TB) were not informed that they were going to assess the same images a second time (in a different order). The images were assessed independently by each investigator, who were blinded to the clinical profiles of the patients, including the endoscopic management of the lesions as well as the histological analysis.

Statistics

All data was collected with predesigned forms. Both SPSS 20.0 (SPSS, Inc., Chicago, IL) and R (version 2.15.1, <http://www.r-project.org/>) were used for statistical analyses. Reproducibility of interpretations was tested by Cohen's kappa coefficient (κ) for inter-observer variability (between two observers, described as median kappa and range) and by Fleiss's kappa coefficient for general inter-observer variability (between ten independent observers, described as kappa and p-value). In general, a kappa value $\kappa > 0.81$ is regarded excellent, $0.80 > \kappa > 0.61$ is substantial, $0.60 > \kappa > 0.41$ is moderate, $0.40 < \kappa < 0.60$ is average, and $\kappa < 0.40$ is poor.¹⁶ Inter-observer agreement was assessed as agreement between readings of two or more observers, whereas intra-observer agreement was assessed as agreement between first and second readings of the same observer. For the inter-observer agreement, only readings from the first round of image assessment were used. Inter- and intra-observer agreement were analysed for all Kudo pit pattern possibilities (6 options), for three groups of Kudo pit pattern possibilities (I and II vs. III_L and IV vs. III_S and V, 3 options), for two groups of Kudo pit pattern possibilities (I, II vs. III_L, III_S, IV and V, 2 options), and finally for the overall assessment of the neoplastic character of the lesion (non-neoplastic vs. neoplastic, 2 options).

Inter- and intra-observer agreement were first analysed for all images together, and further separately for the different endoscopic techniques (CE vs. NBI). Mann Whitney U test was performed to assess differences in inter- and intra-observer agreement between different endoscopic techniques. Wilcoxon signed rank test was performed to assess differences between groups of endoscopic scoring.

Finally, agreement between endoscopic (Kudo pit pattern, overall endoscopic assessment) and histological assessment (non-neoplastic vs. neoplastic) was assessed using Chi-square statistics. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy, described as median and IQR were also computed.

All hypotheses were tested at the 5% level of significance, p value < 0.05 was used for significance threshold.

RESULTS

Accuracy to predict neoplastic lesions

Thirteen out of fifty lesions were neoplastic on histology. As shown in Table 2, median (range) sensitivity, specificity, NPV and PPV for diagnosing neoplasia based on the presence of pit pattern III_L, III_S, IV or V was 77% (54-85%), 68% (51-84%), 88% (84-94%) and 46% (36-61%), respectively. Similarly, based on the overall endoscopic appearance of the lesion, sensitivity, specificity, NPV and PPV was 77% (31-100%), 69% (43-92%), 90% (79-100%), 48% (37-67%), respectively. The NPV was comparable for lesions detected by CE or NBI (p=0.971 for I and II, vs. III_L, III_S, IV or V; and p=0.739 for non-neoplastic vs. neoplastic appearance).

Level of certainty

During the first assessment, the endoscopists were sure or very sure about the neoplastic nature of the lesion of interest in 62% (range 34%-90%) of the cases. The overall diagnostic accuracy was significantly better when endoscopists had a high level of confidence (77% versus 21% (p<0.001)). This was mainly due to an improvement of the PPV (35% versus 58%). No influence was seen on the NPV (Table 3).

1

2 **Inter- and intra-observer agreement**

3 The overall inter-observer agreement for the first assessment of the images is depicted in Table 4.

4 Inter-observer agreement was significantly worse for the complete Kudo pit pattern assessment with

5 6 different options compared to all other endoscopic groups (Kudo I and II vs. III_L and IV vs. III_S and V;

6 Kudo I, II vs. III_L, III_S, IV and V; and overall assessment of the neoplastic character of the lesion; all

7 $p < 0.001$). The overall inter-observer agreement for any pit pattern assessment with six options was

8 only fair ($\kappa = 0.282$), but was significantly better with CE ($\kappa = 0.322$ for CE vs. $\kappa = 0.224$ for NBI, $p = 0.001$).

9 The agreement for differentiation between non-neoplastic (I, II) and neoplastic pit patterns (III_L, III_S,

10 IV, V) was moderate ($\kappa = 0.587$), but significantly better and substantial for NBI ($\kappa = 0.653$ for NBI vs. κ

11 0.495 for CE, $p < 0.001$). Similarly, the overall endoscopic assessment of the neoplastic character of the

12 lesion was better for NBI ($\kappa = 0.564$) in comparison to HD-CE ($\kappa = 0.493$) ($p < 0.001$).

13 The median (range) intra-observer agreement between the two assessments is depicted in Table 5.

14 Although the range in intra-observer agreement among the different observers is wide, intra-observer

15 agreement is moderate for the overall pit pattern assessment, and substantial for differentiating

16 neoplastic pit patterns from non-neoplastic pit patterns. Intra-observer agreement was significantly

17 worse for the complete Kudo pit pattern assessment with 6 different options compared to all other

18 combinations (Kudo I and II vs. III_L and IV vs. III_S and V; and Kudo I, II vs. III_L, III_S, IV and V; both $p = 0.005$).

19 The intra-observer agreement was similar for images derived from endoscopic procedures with HD-CE

20 and NBI (all $p > 0.05$).

21

22

DISCUSSION

This is the first pilot study assessing the diagnostic accuracy and interobserver agreement of both non-magnified HD-CE and NBI of visible lesions detected during UC surveillance. This study assessed the applicability of Kudo's pit pattern in UC lesions and proposes a simplified dichotomous classification of neoplastic versus non-neoplastic patterns. The presence of a pit pattern I or II has a good NPV to rule out dysplasia, with an acceptable overall sensitivity of 77%. We have shown that the inter-observer agreement for differentiating neoplastic from non-neoplastic pit patterns in patients with long standing UC, is moderate to substantial for endoscopists with extensive experience in CE and NBI. We also found that the inter-observer agreement for differentiation between non-neoplastic and neoplastic pit patterns is significantly better for NBI in comparison to HD-CE, although the agreement for the individual pit patterns is better for HD-CE. We believe our data support the development of a prospective trial in consecutive patients referred for UC surveillance, to validate the application of this simplified dichotomous classification of the pit pattern using NBI and non-magnified HD-CE.

In general, little is known about the diagnostic accuracy of non-magnified HD-CE in predicting histology of lesions detected during UC surveillance endoscopy. One study showed a high diagnostic accuracy of magnified standard resolution CE in 87 patients undergoing CE with methylene blue. The prediction of neoplasia was based on the assessment of a type I or II pit pattern for non-neoplastic changes and type III-V for the prediction of neoplastic changes. In the latter study the single-operator sensitivity and specificity were 93% both, with a PPV of 83%, and NPV of 98%.⁵ Similarly, the original Kudo pit pattern was designed for magnifying endoscopy using a combination of indigo carmine and cresyl violet. It is interesting to look back at this original publication and realize that this "gold standard" had 81.5% correlation with stereomicroscopic assessment.¹⁵ Interestingly, no hyperplastic polyps were included in the Kudo paper, but only 2.8% of the adenomas exhibited a type I/II non-neoplastic pattern. A more recent image based study assessed the diagnostic accuracy of low-grade dysplasia in patients with inflammatory bowel disease based on non-magnified HD-CE¹⁷. These authors had 30 (13 low grade

dysplasia) lesions assessed by 17 endoscopists. In contrast to our findings, the interobserver agreement to differentiate between neoplastic and non-neoplastic lesions was only fair with a Kappa of 0.24. This may seem contradictory but the endpoints of assessment were different from our study. Indeed, in the study from Wanders et al, endoscopists had to indicate if the lesion was neoplastic or not. In our study, we first asked to assess the pit pattern, keeping in mind that pattern I and II were non-neoplastic, and at the end the lesions needed to be classified as neoplastic or non-neoplastic. The diagnostic accuracy of identifying neoplastic lesions and pit pattern I/II versus others was almost identical, indicating that in our study the latter was the main driving force to classify a lesion, and may possibly increase the interobserver agreement. In addition, in our trial we also included more advanced lesions.

It is important to realize that pit pattern assessment may be challenging in the setting of UC, with NBI. Matsumoto *et al* already described that unlike the fact that type III_L and IV pit pattern are predictive of neoplasia in screening colonoscopy, they could not find that a villous pattern in UC patients predicted neoplasia. When using NBI, it seems to be that protruding lesions and a tortuous pattern are suggestive of neoplasia.¹⁸ Case reports also show the challenge of differentiating inflammatory polyps exhibiting a type IV pattern.¹⁹ All studies assessing the diagnostic yield of CE for colitis associated neoplasia, were performed in the absence of active inflammation. This is a prerequisite for good quality screening and to enable optical diagnosis based on the pit pattern.

Experience is an important issue in optical diagnosis and pit pattern assessment. For standard polyp pit pattern assessment with magnifying endoscopy, the inter- and intra-observer kappa values are 0.716 and 0.810, respectively, for experienced Japanese endoscopists for perfectly selected images. In the latter study, the inter- and intra-observer agreement for predicting histology was better than for pit pattern analysis alone, which is similar to our findings.²⁰ The learning curve for optical diagnosis in UC is unknown. A learning effect after formal training has been described for pit and vessel pattern analysis using CE and NBI with low and high magnification. Pre-training assessment of 220 high quality

1 images and a formal one hour training on pit pattern and vessel pattern classification improves the
2 inter-observer Kappa value for endoscopists with some experience, but does not bring this to the level
3 of highly experienced endoscopists. Moreover, endoscopists without any experience in optical
4 diagnosis do not perform to the level of more experienced endoscopists, even after a formal training
5 ²¹. This is also echoed by the study from Wanders et al ¹⁷ showing a lower diagnostic accuracy for non-
6 experts in predicting low grade dysplasia.

7 In this study, all endoscopists had been involved in previous imaging studies, but with different levels
8 of experience. In addition, we used non-magnified images and we used the pit pattern assessment in
9 the setting of UC. These factors explain why the overall Kappa value is comparable to endoscopists
10 with less experience assessing non-magnified images. Nonetheless, the findings of our pilot study are
11 valuable since the use of non-magnifying endoscopy represents better real life endoscopy in general
12 secondary care units. In addition, we have shown that using a dichotomous assessment of the pit
13 pattern (Type I and II, versus the rest for prediction of neoplasia), we obtained a relatively good
14 sensitivity for diagnosing neoplastic lesions, even without magnifying endoscopy, which is comparable
15 to the diagnostic accuracy obtained by highly experienced endoscopists for assessing standard polyps
16 ²¹. We also included a clinically realistic number of neoplastic lesions in the image data set (26%), which
17 is comparable to previously reported neoplasia detection rate in UC.^{5,10,11}

18 The level of certainty about the optical diagnosis needs to be considered. For instance, for the NICE
19 classification, a high level of confidence is necessary to make an optical NBI based diagnosis.²² Similarly,
20 we found a significant improvement in diagnostic accuracy when endoscopists felt confident about the
21 diagnosis. Although we did not assess this in the study, one can speculate that the level of confidence
22 will mainly be determined by the quality of the image and to some extent the presence of
23 inflammation. We opted not to use the NICE classification because it has not been validated for UC
24 lesions. Unlike the fact that NBI has mainly been attributed to highlight vascular structures, we did

show however that non-magnifying NBI of high quality still image allows to distinguish between neoplastic and non-neoplastic pit patterns with a high interobserver and intra-observer agreement.

The limitation of our study is that all image assessment was performed off-site and on selected still images by experts. It is not clear how this translates to assessment of videos or assessment during endoscopy by less experienced endoscopists. Although the investigator who selected the images for this study was otherwise not involved in either the original study, nor in the interobserver study, a selection bias is still possible. Recall bias was probably limited due to the interval of at least 10 weeks in between the two assessments. In addition, one of the assessors (RB) performed the original endoscopy and 2 assessors (RB and TB) were aware that the same set of images were being used during the second evaluation. However a sensitivity analysis excluding these two assessors did not change the overall findings of the study. Also, size of the lesions was not taken into account as a possible confounding factor, but in case of UC, even inflammatory polyps can be very large. Finally, due to the relatively small sample size of the images, our findings need to be confirmed prospectively, preferably in a real-time scenario during ongoing surveillance.

What can we learn from this pilot study for general endoscopy practice? We clearly showed that inter-observer agreement as well as diagnostic accuracy of optical diagnosis in the setting of non-magnifying HD-CE or NBI for assessment of UC associated neoplasia is not perfect in a group of endoscopists with mixed levels of experience. It can be anticipated, based on available literature, that endoscopists without any experience in optical diagnosis are unable to make the right call about pit pattern and final histological diagnosis¹⁷. This is already the case for standard polyp assessment,²¹ and most likely pertains in a setting of UC. Training programs should be provided, and in the absence of any experience, endoscopists performing UC surveillance should biopsy any visible lesion for proper histological diagnosis and to get feedback on their own findings. As for the more experienced endoscopists, it is reassuring to realize that the inter-observer agreement is moderate to substantial to differentiate a non-neoplastic from a neoplastic pit pattern even without magnification, and that

1 the diagnostic accuracy for this dichotomous assessment approaches expert levels. Nevertheless, the
2 system is not perfect and therefore it is advised to take biopsies of any lesion, unless the diagnosis of
3 a very typical pseudopolyp with pit pattern type I, or a clearly identifiable small hyperplastic lesion is
4 made with a high level of confidence. In case of larger lesions and especially when the level of
5 confidence is low, biopsies should be taken for pathological confirmation.

6 In conclusion, a dichotomous pit pattern assessment after non-magnifying CE or NBI has a moderate
7 to substantial inter-observer agreement and a good diagnostic accuracy for differentiating neoplastic
8 from non-neoplastic lesions in UC surveillance. Diagnostic accuracy improves significantly with the
9 level of confidence. However, even with experience in optical diagnosis, taking biopsies of lesions that
10 are not typical pseudopolyps or hyperplastic polyps remains necessary. The findings of this pilot study
11 need to be confirmed in a prospective study with a real-time scenario during ongoing UC surveillance.

12 Acknowledgements

13 RB and MF Marc Ferrante are Senior Clinical Investigators for the Research Foundation–Flanders
14 (FWO), Belgium.

REFERENCES

1. Eaden JA, Abrams KR, Mayberry JF. The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut* 2001;48(4):526–35.
2. Jess T, Loftus E V, Velayos FS, Harmsen WS, Zinsmeister AR, Smyrk TC, et al. Risk of intestinal cancer in inflammatory bowel disease: a population-based study from olmsted county, Minnesota. *Gastroenterology*. 2006;130(4):1039–46.
3. Jess T, Simonsen J, Jørgensen KT, Pedersen BV, Nielsen NM, Frisch M. Decreasing risk of colorectal cancer in patients with inflammatory bowel disease over 30 years. *Gastroenterology*. 2012 Aug;143(2):375-81-4.
4. Rutter M, Saunders B, Wilkinson K, Rumbles S, Schofield G, Kamm M, et al. Severity of inflammation is a risk factor for colorectal neoplasia in ulcerative colitis. *Gastroenterology*. 2004 Feb;126(2):451–9.
5. Kiesslich R, Fritsch J, Holtmann M, Koehler HH, Stolte M, Kanzler S, et al. Methylene blue-aided chromoendoscopy for the detection of intraepithelial neoplasia and colon cancer in ulcerative colitis. *Gastroenterology*. 2003 Apr;124(4):880–8.
6. Marion J, Waye J, Present D. Chromoendoscopy-targeted biopsies are superior to standard colonoscopic surveillance for detecting dysplasia in inflammatory bowel disease patients: a prospective. *Am J Gastroenterol*. 2008;103(9):2342–9.
7. Subramanian V, Mannath J, Ragunath K, Hawkey CJ. Meta-analysis: the diagnostic yield of chromoendoscopy for detecting dysplasia in patients with colonic inflammatory bowel disease. *Aliment Pharmacol Ther*. 2011 Feb;33(3):304–12.
8. Cairns SR, Scholefield JH, Steele RJ, Dunlop MG, Thomas HJW, Evans GD, et al. Guidelines for colorectal cancer screening and surveillance in moderate and high risk groups (update from 2002). *Gut*. 2010 May ;59(5):666–89.
9. Kamiński MF, Hassan C, Bisschops R, Pohl J, Pellisé M, Dekker E, et al. Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. *Endoscopy*. 2014 May;46(5):435–49.
10. Pellisé M, López-Cerón M, Rodríguez de Miguel C, Jimeno M, Zabalza M, Ricart E, et al. Narrow-band imaging as an alternative to chromoendoscopy for the detection of dysplasia in long-standing inflammatory bowel disease: a prospective, randomized, crossover study. *Gastrointest Endosc*; 2011;74(4):840–8.
11. Ignjatovic A, East JE, Subramanian V, Suzuki N, Guenther T, Palmer N, et al. Narrow band imaging for detection of dysplasia in colitis: a randomized controlled trial. *Am J Gastroenterol*; 2012 Jun ;107(6):885–90.
12. R. Bisschops, T. Bessissow, JA Josph, F. Baert, M. Ferrante, V. Ballet, H. Willekens, I. Demedts, K. Geboes, G. De Hertogh, S. Vermeire, P. Rutgeerts, G. Van Assche. Chromo-Endoscopy Versus Narrow Band Imaging in Ulcerative Colitis: A Prospective Randomized Controlled Trial. *Gut*. 2017 Jul 11. pii: gutjnl-2016-313213. doi: 10.1136/gutjnl-2016-313213. [Epub ahead of print]
13. Laine L, Kaltenbach T, Barkun A, McQuaid KR, Subramanian V, Soetikno R. SCENIC international consensus statement on surveillance and management of dysplasia in inflammatory bowel disease. *Gastroenterology*. 2015;148(3):639–651.e28.

- 1 14. Schlemper RJ, Riddell RH, Kato Y, Borchard F, Cooper HS, Dawsey SM, et al. The Vienna
2 classification of gastrointestinal epithelial neoplasia. *Gut*. 2000;47(2):251–5.
- 3 15. Kudo S, Tamura S, Nakajima T, Yamano H, Kusaka H, Watanabe H. Diagnosis of colorectal
4 tumorous lesions by magnifying endoscopy. *Gastrointest Endosc*. 1996 Jul ;44(1):8–14.
- 5 16. Altman D. *Practical Statistics for Medical Research*. Chapman & Hall C, editor. London; 1990.
6 403-408 p.
- 7 17. Wanders LK, Mooiweer E, Wang J, Bisschops R, Offerhaus GJ, Siersema PD, et al. Low
8 interobserver agreement among endoscopists in differentiating dysplastic from non-dysplastic
9 lesions during inflammatory bowel disease colitis surveillance. *Scand J Gastroenterol*.
10 2015;50(8):1011–7.
- 11 18. Matsumoto T, Kudo T, Jo Y, Esaki M, Yao T, Iida M. Magnifying colonoscopy with narrow band
12 imaging system for the diagnosis of dysplasia in ulcerative colitis: a pilot study. *Gastrointest*
13 *Endosc* 2007;66(5):957–65.
- 14 19. East J, Suzuki N, Herbay A Von, Saunders B. Narrow band imaging with magnification for
15 dysplasia detection and pit pattern assessment in ulcerative colitis surveillance: a case with
16 multiple dysplasia associated. *Gut*.;55(10):1432–5.
- 17 20. Huang Q, Fukami N, Kashida H, Takeuchi T, Kogure E, Kurahashi T, et al. Interobserver and
18 intra-observer consistency in the endoscopic assessment of colonic pit patterns. *Gastrointest*
19 *Endosc*. 2004;60(4):520–6.
- 20 21. Higashi R, Uraoka T, Kato J, Kuwaki K, Ishikawa S, Saito Y, et al. Diagnostic accuracy of narrow-
21 band imaging and pit pattern analysis significantly improved for less-experienced
22 endoscopists after an expanded training program. *Gastrointest Endosc*. 2010;72(1):127–35.
- 23 22. Hayashi N, Tanaka S, Hewett DG, Kaltenbach TR, Sano Y, Ponchon T, et al. Endoscopic
24 prediction of deep submucosal invasive carcinoma: validation of the narrow-band imaging
25 international colorectal endoscopic (NICE) classification. *Gastrointest Endosc*. 2013
26 Oct;78(4):625–32.

27

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

Legend to Figures

Figure 1. Two images originating from the PowerPoint file for the first round. Image 1.22 was initially regarded as a Kudo I pit pattern during endoscopy with narrow band imaging, while final histology showed a hyperplastic polyp (A). Images 1.43 was initially regarded as a Kudo III_L pit pattern during endoscopy with chromo-endoscopy, while final histology demonstrated low grade dysplasia (B).

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10

Table 1. Distribution of morphology of the lesions according to the Scenic consensus modified Paris classification.

	Neoplastic lesions	Non-neoplastic lesions
Polypoid		
Pedunculated (Ip)	0	2
Sessile (Is)	2	9
Nonpolypoid		
Flat		
IIa	11	17
IIb	0	8
Depressed (IIc)	0	1

Table 2. Accuracy to predict histologic proven neoplastic lesion

	Sensitivity	Specificity	NPV	PPV	Accuracy
Overall (n=50)					
I-II vs. III _L -IV-III _S -V	77% (54%-85%)	68% (51%-84%)	88% (84%-94%)	46% (36%-61%)	70% (58%-82%)
Non-neoplastic vs. neoplastic	77% (31%-100%)	69% (43%-92%)	90% (79%-100%)	48% (37%-67%)	72% (58%-84%)
CE (n=24)					
I-II vs. III _L -IV-III _S -V	88% (63%-100%)	63% (44%-81%)	89% (78%-100%)	57% (40%-73%)	73% (54%-88%)
Non-neoplastic vs. neoplastic	88% (25%-100%)	59% (38%-94%)	91% (71%-100%)	55% (40%-78%)	71% (54%-88%)
NBI (n=26)					
I-II vs. III _L -IV-III _S -V p versus CE	60% (40%-80%) p <0.001	74% (57%-86%) p 0.280	89% (86%-92%) p 0.739	35% (25%-43%) p <0.001	71% (58%-77%) p 0.481
Non-neoplastic vs. neoplastic p versus CE	60% (40%-100%) p 0.132	76% (48%-90%) p 0.315	89% (86%-100%) p 0.971	38% (27%-78%) p 0.015	73% (58%-81%) p 0.481

Diagnostic performance values expressed as median with (range)

CE: chromo-endoscopy with methylene blue; NBI: narrow band imaging; NPV: negative predictive value; PPV: positive predictive value

The diagnostic accuracy is first calculated based on the dichotomous model pit pattern I-II versus other pit patterns and then based on the overall impression of the lesion as assessed by the endoscopist (non-neoplastic vs neoplastic).

1

2 **Table 3. Diagnostic accuracy for neoplasia per level of confidence of the observer**

3

	Not sure (1-3)	Sure (4-5)	p-value
Median sensitivity (range)	73% (0% - 100%)	80% (43% - 100%)	0.436
Median specificity (range)	66% (25% - 100%)	73% (56% - 89%)	0.353
Median NPV (range)	90% (50% - 100%)	89% (81% - 100%)	0.684
Median PPV (range)	35% (0% - 100%)	58% (38% - 70%)	0.052
Median accuracy (range)	21% (4% - 50%)	77% (62% - 84%)	<0.001

4

5 NPV: negative predictive value; PPV: positive predictive value

6

Table 4. Inter-observer agreement

	Overall (n=50)	CE (n=24)	NBI (n=26)
Kudo pit pattern	K = 0.282 p < 0.001	K = 0.322 p < 0.001	K = 0.224 p < 0.001
I-II vs. III_L-IV vs. III_S-V	K = 0.497 p < 0.001	K = 0.462 p < 0.001	K = 0.492 p < 0.001
I-II vs. III_L-IV-III_S-V	K = 0.587 p < 0.001	K = 0.495 p < 0.001	K = 0.653 p < 0.001
Non-neoplastic vs. neoplastic	K = 0.493 p < 0.001	K = 0.394 p < 0.001	K = 0.564 p < 0.001

The p-value was generated by calculating the variance of kappa and deriving a z statistics

CE: chromo-endoscopy with methylene blue; NBI: narrow band imaging

Table 5. Intra-observer agreement

	Overall (n=50)	CE (n=24)	NBI (n=26)
Kudo pit pattern	K = 0.507 (0.270 – 0.565)	K = 0.464 (0.249 – 0.671)	K = 0.412 (0.152 – 0.776)
I-II vs. III_L-IV vs. III_S-V	K = 0.642 (0.373 – 0.862)	K = 0.527 (0.249 – 0.934)	K = 0.686 (0.226 – 0.891)
I-II vs. III_L-IV-III_S-V	K = 0.685 (0.436 – 0.960)	K = 0.587 (0.249 – 1.000)	K = 0.756 (0.350 – 0.906)
Non-neoplastic vs. neoplastic	K = 0.663 (0.160 – 0.960)	K = 0.577 (0.216 – 1.000)	K = 0.738 (0.025 – 1.000)

Intra-observer agreement: Median Kappa with (range)

CE: chromo-endoscopy with methylene blue; NBI: narrow band imaging