Advances in colonoscopic imaging and the approach to dysplasia in IBD
Wanders, L.K.

Citation for published version (APA):
Wanders, L. K. (2016). Advances in colonoscopic imaging and the approach to dysplasia in IBD
CHAPTER 3

Low inter-observer agreement among endoscopists in differentiating dysplastic from non-dysplastic lesions during IBD colitis surveillance

L.K. WANDERS
E. MOOIWEER
J. WANG
R. BISSCHOPS
G.J. OFFERHAUS
P.D. SIERSEMA
G.R. D’HAENS
B. OLDENBURG
E. DEKKER

Scand J Gastroenterol. 2015 Aug;50(8):1011-7
ABSTRACT

Introduction
During endoscopic surveillance in patients with longstanding colitis a variety of lesions can be encountered. Differentiation between dysplastic and non-dysplastic lesions can be challenging. The accuracy of visual endoscopic differentiation and inter-observer agreement (IOA) has never been objectified.

Methods
We assessed the accuracy of expert and non-expert endoscopists in differentiating (low-grade) dysplastic from non-dysplastic lesions and the IOA among and between them. An online questionnaire was constructed containing 30 cases including a short medical history and an endoscopic image of a lesion found during surveillance employing chromoendoscopy.

Results
A total of 17 endoscopists, 8 experts and 9 non-experts, assessed all 30 cases. The overall sensitivity and specificity for correctly identifying dysplasia were 73.8% (95% CI 62.1-85.4) and 53.8% (95% CI 42.6-64.7), respectively. Experts showed a sensitivity of 76.0% (95% CI 63.3-88.6) versus 71.8% (95% CI 58.5-85.1, p=0.434) for non-experts, the specificity 61.0% (95% CI 49.3-72.7) versus 47.1% (95% CI 34.6-59.5, p=0.008). The overall IOA in differentiating between dysplastic and non-dysplastic lesions was fair 0.24 (95% CI 0.21-0.27); for experts 0.28 (95% CI 0.21-0.35) and for non-experts 0.22 (95% CI 0.17-0.28). The overall IOA for differentiating between subtypes was fair 0.21 (95% CI 0.20-0.22); for experts 0.19 (95% CI 0.16-0.22) and non-expert 0.23 (95% CI 0.20-0.26).

Conclusion
In this image-based study, both expert and non-expert endoscopists cannot reliably differentiate between dysplastic and non-dysplastic lesions. This emphasizes that all lesions encountered during colitis surveillance with a slight suspicion of containing dysplasia should be removed and sent for pathological assessment.
INTRODUCTION

Patients with inflammatory bowel disease (IBD) are at increased risk of developing colorectal cancer (CRC).\(^1\)\(^-\)\(^5\) Since the risk of colitis-associated CRC increases with a longer disease duration and greater disease extent\(^6\)\(^,\)\(^7\), current guidelines recommend colonoscopic surveillance in all IBD patients with longstanding extensive colitis.\(^8\)\(^-\)\(^10\)

During surveillance colonoscopy in IBD patients, a variety of dysplastic and non-dysplastic lesions can be found. Dysplastic lesions can be classified as either sporadic adenomas, i.e. not related to the inflammation, or colitis-associated dysplasia, also known as dysplasia associated lesion or mass (DALM). Differentiation between these entities is challenging, as there are no distinct histological features to confirm the diagnosis. In the past, the risk of synchronous and metachronous colorectal cancer was considered to be very high in patients with lesions classified as DALM and therefore, prophylactic proctocolectomy was advocated\(^11\). Nowadays there is ongoing debate whether endoscopic resection is justified in some of these cases.

Non-dysplastic lesions, comprising post-inflammatory polyps, lesions caused by scarring or active inflammation, and common non-dysplastic lesions such as hyperplastic polyps can be found as well. This wide variety of lesions renders differentiation between dysplastic and non-dysplastic lesions challenging. Several studies have demonstrated that pan-chromoendoscopy (pCE), employing methylene blue or indigo carmine, improves the detection of dysplastic lesions during IBD surveillance colonoscopies\(^12\)\(^-\)\(^14\), and therefore current guidelines recommend this technique as the preferred method.\(^9\)\(^-\)\(^10\) In addition, pCE could also aid in the endoscopic differentiation of lesions found during colonoscopy.\(^15\)

In the current study we aimed to investigate the inter-observer accuracy, agreement and variability between expert and non-expert endoscopists in differentiating dysplastic from non-dysplastic lesions encountered in patients with longstanding colitis undergoing pCE surveillance.
METHODS

An online questionnaire (LimeSurvey, www.limesurvey.com) was constructed containing images of 30 lesions encountered during CRC surveillance in IBD patients. High quality images of lesions were selected from a database containing patients with longstanding ulcerative colitis (UC) and Crohn’s disease (CD), enrolled in a surveillance program in two academic centres from 2009 to 2013. Cases were excluded when disease activity was present during surveillance colonoscopy, rated by the performing endoscopist. In all cases, Olympus CF-160 and CF-180 colonoscopes were used and pCE with either methylene blue 0.1% or indigo-carmine 0.3% was employed. For every case included in the study, biopsies sampled from each lesion were available and histological diagnosis served as the golden standard. One endoscopic image of each lesion was uploaded into the questionnaire without any form of post-processing besides removing the name of the patient from the image and highlighting the lesion with an arrow. In total, 13 lesions containing low-grade dysplasia (LGD) and 17 non-dysplastic lesions were selected. One expert GI-pathologist reassessed all dysplastic lesions available according to Vienna classification in order to limit false positive findings.¹⁶

The cases were numbered, and for each case, two slides were prepared. The first one listed a standardized short medical history (age, gender, type and duration of IBD, disease extent, prior dysplasia, and concomitant diagnosis of primary sclerosing cholangitis) and an endoscopic image of the lesion, the other slide showed the pathological diagnosis.

Participating endoscopists

Twelve consultant gastroenterologists and five fellows in training completed the questionnaire. Eight gastroenterologists were classified as “expert” as they were working in a referral centre and had performed at least 50 IBD surveillance colonoscopies with pCE. The non-expert group consisted of five fellows in training and four endoscopists working in a general hospital. No specific training was given prior to the questionnaire, although a reference to the current IBD-surveillance guideline ⁹ and practical instructions on the questionnaire were provided.
Assessment
The participants were asked to classify each lesion into seven categories: sporadic adenoma, adenoma-like DLM, non-adenoma-like DLM, inflammation, post-inflammatory polyp, hyperplastic polyp or normal mucosa. The first three types of lesions (sporadic adenoma, adenoma-like DLM and non-adenoma-like DLM) were classified as dysplastic lesions; the remaining four were combined as non-dysplastic. The confidence of each endoscopic diagnosis was assessed by asking the participants to rate the certainty of each diagnosis from 1 to 5, where 1 was 'very uncertain' and 5 'very certain'.

After the first assessment of each lesion, the histopathological diagnosis of the lesion was provided to the participant. As the sub-differentiation of especially the dysplastic lesions is dependent on the endoscopist, participants were asked to classify all the lesions again after histopathology.

Statistical analysis
The sensitivity and specificity for differentiating dysplastic from non-dysplastic lesions were calculated using histopathology of each lesion as reference standard. The sensitivity, specificity, diagnostic accuracy, negative predictive value (NPV) and positive predictive value (PPV) were calculated and compared between experts and non-experts using generalized estimating equation (GEE) with exchangeable correlation structure. GEE was also used for the analysis of the effect of confidence on test accuracy.

The overall inter-observer agreement for the differentiation between dysplastic and non-dysplastic lesions as well as for each subcategory was calculated using Fleiss Kappa, which measures the level of agreement beyond chance. Interpretation of the Kappa values were done according Landis and Koch. The inter-observer agreement was also calculated separately for experts and non-experts and compared with Wald statistic by assuming the Kappa of experts and non-experts had a correlation of 0.3 since they are observing lesions from the same patient group. The post-pathology differentiation between sporadic adenoma, adenoma-like and non-adenoma-like DLM was not compared to a gold standard since there is no uniform definition for these three classifications. The agreement on the differentiation between adenoma-like and non-adenoma-like DLM was calculated based on the classification given
after the pathological diagnosis of the lesion. The two dependent kappa’s were compared according to Donner.  

RESULTS

Participants were able to correctly classify lesions as dysplastic with a sensitivity of 73.8% (95% CI 66.1-85.4), a specificity of 53.8% (95% CI 42.6-64.7) and an accuracy of 62.4% (95% CI 53.8-71.1), shown in table 1. The overall inter-observer agreement for the differentiation between dysplastic and non-dysplastic lesions was fair with a Kappa of 0.24 (95% CI 0.21-0.27).

A dysplastic lesion that often incorrectly assessed as non-dysplastic are shown in figure 1, a non-dysplastic lesion that was often assessed as dysplastic are shown in figure 2.

**Comparison between experts and non-experts**

The sensitivity for identifying lesions containing dysplasia was 76.0% for experts, which was not significantly different from the non-experts (71.8%, p=0.50), but experts had a significantly higher specificity (61.0% versus 47.1% respectively, p=0.008, table 1).
Table 1 Inter-observer agreement and diagnostic accuracy for the differentiation between dysplastic and non-dysplastic lesions

<table>
<thead>
<tr>
<th></th>
<th>Overall (n=17)</th>
<th>Experts (n=8)</th>
<th>Non-experts (n=9)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-observer agreement</td>
<td>0.24 (0.21 – 0.27)</td>
<td>0.28 (0.21 – 0.35)</td>
<td>0.22 (0.17 – 0.28)</td>
<td>0.163</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>73.8 (62.1 – 85.4)</td>
<td>76.0 (63.3 – 88.6)</td>
<td>71.8 (58.5 – 85.1)</td>
<td>0.434</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>53.8 (42.6 – 64.7)</td>
<td>61.0 (49.3 – 72.7)</td>
<td>47.1 (34.6 – 59.5)</td>
<td>0.008</td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>62.4 (53.6 – 71.1)</td>
<td>67.5 (58.5 – 76.5)</td>
<td>57.8 (47.6 – 67.9)</td>
<td>0.010</td>
</tr>
<tr>
<td>NPV (%)</td>
<td>73.0 (68.1 – 77.9)</td>
<td>76.8 (72.2 – 81.4)</td>
<td>68.5 (60.9 – 76.1)</td>
<td>0.064</td>
</tr>
<tr>
<td>PPV (%)</td>
<td>54.3 (49.6 – 59.1)</td>
<td>59.8 (51.5 – 68.2)</td>
<td>50.8 (46.0 – 55.7)</td>
<td>0.109</td>
</tr>
</tbody>
</table>

NPV = negative predictive value, PPV = positive predictive value

The inter-observer agreement for the differentiation between dysplastic and non-dysplastic lesions was fair, for experts as well as non-experts, 0.28 (95% CI 0.21-0.35) and 0.22 (95% CI 0.17-0.28) respectively (p=0.163).

The overall inter-observer agreement for each subcategory is shown in table 2. The agreement was highest for the classifications post-inflammatory polyp, hyperplastic polyp and sporadic adenoma (all κ=0.26) and lowest for non-adenoma-like DALM (κ=0.12). Experts showed a significantly higher agreement for the category ‘post-inflammatory polyp,’ while non-experts showed a significantly higher agreement for the categories normal mucosa and ‘non-adenoma-like DALM’ (table 2).

Table 2 Inter-observer agreement before pathological diagnosis was given

<table>
<thead>
<tr>
<th></th>
<th>Overall (n=17)</th>
<th>Experts (n=8)</th>
<th>Non-experts (n=9)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.21 (0.20 – 0.22)</td>
<td>0.19 (0.16 – 0.22)</td>
<td>0.23 (0.20 – 0.26)</td>
<td>0.033</td>
</tr>
<tr>
<td>Normal mucosa</td>
<td>0.20 (0.17 – 0.23)</td>
<td>0.08 (0.01 – 0.14)</td>
<td>0.28 (0.22 – 0.34)</td>
<td>0.000</td>
</tr>
<tr>
<td>Inflammation</td>
<td>0.19 (0.16 – 0.22)</td>
<td>0.18 (0.11 – 0.24)</td>
<td>0.15 (0.09 – 0.21)</td>
<td>0.516</td>
</tr>
<tr>
<td>Post-inflammatory polyp</td>
<td>0.26 (0.23 – 0.29)</td>
<td>0.36 (0.29 – 0.43)</td>
<td>0.19 (0.13 – 0.25)</td>
<td>0.000</td>
</tr>
<tr>
<td>Hyperplastic polyp</td>
<td>0.26 (0.23 – 0.29)</td>
<td>0.27 (0.20 – 0.34)</td>
<td>0.27 (0.21 – 0.32)</td>
<td>0.876</td>
</tr>
<tr>
<td>Sporadic adenoma</td>
<td>0.26 (0.23 – 0.29)</td>
<td>0.22 (0.15 – 0.29)</td>
<td>0.28 (0.22 – 0.34)</td>
<td>0.126</td>
</tr>
<tr>
<td>Adenoma-like DALM</td>
<td>0.17 (0.14 – 0.20)</td>
<td>0.14 (0.07 – 0.20)</td>
<td>0.20 (0.14 – 0.26)</td>
<td>0.113</td>
</tr>
<tr>
<td>Non-adenoma-like DALM</td>
<td>0.12 (0.09 – 0.16)</td>
<td>0.10 (0.03 – 0.17)</td>
<td>0.19 (0.13 – 0.25)</td>
<td>0.016</td>
</tr>
</tbody>
</table>

Non-dysplastic: Normal mucosa, inflammation, post-inflammatory polyps, hyperplastic polyp
Dysplastic: Sporadic adenoma, adenoma-like DALM, non-adenoma-like DALM

Confidence level

For every diagnosis the observers were asked for confidence level of their endoscopic classification, on a scale from 1 to 5. The overall confidence level was
3.45 (95% CI 3.33-3.57). For experts this was significantly higher at 3.59 (95% CI 3.46-3.72) than for non-experts (3.33, 95% CI 3.19-3.47) (p=0.001). Of all diagnoses made, 42.9% (219 out of 510) were made with a confidence level 4 and 10.4% (53 out of 510) with a confidence level 5. The relation between the confidence level and accuracy of differentiating between dysplastic and non-dysplastic lesions for both experts and non-experts is shown in figure 3. For the lowest confidence level the accuracy was 36.8% and for the highest confidence level 84.9%. The result of the logistic regression of confidence level on test accuracy showed that for one unit increase in the level of confidence, a 35.7% increase in the odds of the test result being accurate is reached.

Figure 3 Relation between accuracy and confidence level differentiating between dysplastic and non-dysplastic lesions for both experts and non-experts

Classification after histopathologic diagnosis

After providing the histopathology report to the participants, the overall inter-observer agreement for the categories sporadic adenoma, adenoma-like DALM and non-adenoma-like DALM assessed for the 13 dysplastic lesions were 0.37, 0.20 and 0.15 respectively (table 3). Only a significant difference between experts and non-experts was found for the category ‘non-adenoma-like DALM’.
Table 3 Inter-observer agreement for classifying lesions containing dysplasia (n=13) as adenoma-like or non-adenoma like DALM after histopathological diagnosis was provided:

<table>
<thead>
<tr>
<th></th>
<th>Overall (n=17)</th>
<th>Experts (n=8)</th>
<th>Non-experts (n=9)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sporadic adenoma</td>
<td>0.37 (0.33 – 0.42)</td>
<td>0.32 (0.22 – 0.42)</td>
<td>0.41 (0.32 – 0.50)</td>
<td>0.147</td>
</tr>
<tr>
<td>Adenoma-like DALM</td>
<td>0.20 (0.15 – 0.24)</td>
<td>0.17 (0.07 – 0.28)</td>
<td>0.26 (0.17 – 0.35)</td>
<td>0.119</td>
</tr>
<tr>
<td>Non-adenoma-like DALM</td>
<td>0.15 (0.10 – 0.19)</td>
<td>0.03 (-0.08 – 0.13)</td>
<td>0.19 (0.10 – 0.28)</td>
<td>0.007</td>
</tr>
</tbody>
</table>

DISCUSSION

We demonstrate that, during CRC surveillance in IBD patients, expert as well as non-expert endoscopists differentiate dysplastic from non-dysplastic lesions with an acceptable overall sensitivity of 74%. The overall specificity was poor at 54%, however, and experts performed significantly better than the non-experts. The inter-observer agreement for the different subtypes of dysplastic and non-dysplastic lesions was low, for both experts and non-experts.

In a study on IBD-surveillance published in 2007, the ability of endoscopists to distinguish adenoma-like from non-adenoma-like DALMs without the use of pCE was assessed.18 In that study, experts showed a sensitivity of 68% for identifying adenoma-like DALMs, 75% for non-adenoma-like DALMs, and 82% for inflammatory polyps. For non-expert endoscopists accuracy was significantly lower. In our study, we did not assess the sensitivity for differentiation between adenoma-like and non-adenoma-like DALMs, as there is no uniform definition for diagnosis of these lesions. However, results seem comparable and experience seems to play an important role in the accuracy of the optical diagnosis of dysplastic lesions during IBD surveillance colonoscopies.

It is commonly accepted that colitis-associated CRC develops along the inflammation- dysplasia-carcinoma-sequence, which is different from the classical adenoma-carcinoma-sequence that is observed in most patients without colitis.21 Colitis-associated dysplastic lesions are thought to occur multifocally, have a flat architecture, are poorly delineated and are therefore difficult to resect radically during colonoscopy. This type of lesion is mostly referred to as DALM and current surveillance guidelines recommend colectomy if such a lesion is encountered, although there is ongoing debate whether this is the optimal treatment option.12,22 This is in contrast to the situation with sporadic adenomas that are also detected...
in patients without colitis, which can generally be visualized without the aid of chromoendoscopy and can more often safely be resected endoscopically.

A meta-analysis focusing on cancer risk after detection of flat dysplasia in longstanding IBD showed an increased risk of developing CRC, while a recent meta-analysis, in whom polypoid lesions were resected, showed a substantially lower cancer risk. Apparently, the risk of developing CRC after resection of a dysplastic lesion depends on its endoscopic morphology and therefore probably on its resectability.

Therefore, accurate differentiation between colitis-associated dysplasia and sporadic adenomas appears to be relevant for decision-making in the treatment of dysplastic lesions. As there are no distinctive pathological markers available, differentiation is mainly based on the endoscopic image. Endoscopic differentiation between dysplastic and non-dysplastic lesions in patients with colitis is challenging however, as the mucosal and vascular irregularities caused by quiescent or active inflammation can mimic the changes observed in dysplastic lesions. A recent study by Allende and colleagues looked at the correlation between endoscopic diagnosis and pathology and showed a poor correlation. In over half of the cases dysplasia was seen histologically in which the endoscopic diagnosis was negative. Although this study was performed from a pathological perspective, the poor correlation between endoscopy and pathology is comparable to our results and reflects the difficulties in making a diagnosis of lesions found during IBD surveillance colonoscopy solely based on the endoscopic image.

It has been suggested that the application of dye during colonoscopy could increase differentiation of detected lesions by highlighting its surface pattern. The accuracy of the overall differentiation described in our study (overall 62%, expert 68% and non-expert 58%) was similar to the accuracy observed in the study from Farraye where standard white light endoscopy was used (overall 65%, expert 75% and non-expert 56%). Although comparing the results of two image-based studies is not the optimal study approach, this suggests that the additional value of pCE in differentiating lesions encountered during IBD surveillance is smaller than previously perceived. On the other hand, pCE might delineate dysplastic mucosa more clearly, thereby increasing the chance of complete endoscopic removal.
It should be recognized that our study, focusing on endoscopic diagnosis of lesions found during chromoendoscopic IBD surveillance colonoscopies, has limitations. Every image-based inter-observer study is subject to selection bias. All 30 cases contained only one still image, which was manually selected from the endoscopic databases if it represented a characteristic lesion, if the quality of the picture was sufficient and if histology was available. In daily practice, pictures and biopsies are mostly taken in lesions of which the endoscopist is uncertain about the diagnosis. Another limitation is that still images were used in contrast to videos. The latter would have provided the participants with the opportunity to examine the mucosa from different angles and in perspective to the colon. Because histological outcome and endoscopic management was given after each case, this may have affected the assessments of the subsequent cases by creating a short learning curve. We tried to correct for this by presenting the lesions in random order and adding a confidence level to the diagnosis. Obviously, experts have far more experience in performing surveillance using pCE and are more convinced about their diagnosis than non-experts, resulting in significantly higher confidence levels of the endoscopic diagnosis. We showed that accuracy was positively correlated with the confidence level of the endoscopists. Furthermore, our definition of expert endoscopist working in a referral center and who performed at least 50 chromoendoscopies as experts is debatable. Since there is no literature on the actual learning curve of chromoendoscopy, it is unknown whether the endoscopists in our study were indeed experts. Lastly, we have included mainly lesions containing LGD and no HGD lesions have been included, potentially biasing our results. On the other hand, lesions containing HGD in the setting of IBD are rare, and therefore we assume that the present selection of lesions reliably reflects daily practice.

Future studies should focus on real-time differentiation of lesions found during IBD surveillance colonoscopy using chromoendoscopy. Also the learning should be assessed in order to determine the definitions expert and non-expert.

Our study emphasizes the difficulty for both expert and non-expert endoscopists to differentiate dysplastic from non-dysplastic lesions in IBD patients, despite the improved endoscopic techniques. Therefore we recommend to remove lesions entirely in case of doubt or, if this appears to be impossible, to take biopsies for histopathological assessment. Consequently, the management of dysplastic
lesions should mainly be based on its resectability and not on its endoscopic characteristics, as there is no uniformity in the endoscopic classification.

In conclusion, endoscopists, both experts and non-experts, cannot reliably differentiate between dysplastic and non-dysplastic lesions encountered during surveillance in IBD patients. This emphasizes the value of pathological assessment of all potential dysplastic lesions that are encountered during colonoscopy-surveillance in IBD.
REFERENCES

24. Wanders LK, Dekker E, et al. Cancer Risk After Resection of Polypoid Dysplasia in Patients With Long-
3565(13)01088-4.

with inflammatory bowel disease: correlation of pathologic and endoscopic findings. Colorectal Dis
2014 May 19.

26. Farraye FA, Waye JD, et al. Variability in the diagnosis and management of adenoma-like and non-
adenoma-like dysplasia-associated lesions or masses in inflammatory bowel disease: an Internet-based
study. Gastrointest Endosc 2007 Sep;66(3):519-29