Development and validation of confocal endomicroscopy diagnostic criteria for low-grade dysplasia in Barrett's esophagus

di Pietro, Massimiliano; Bertani, Helga; O'Donovan, Maria; Santos, Patricia; Alastal, Hani; Phillips, Richard; Ortiz-Fernandez-Sordo, Jacobo; Iacucci, Marietta; Modolell, Ines; Reggiani Bonetti, Luca; Ragunath, Krish; Wernisch, Lorenz

DOI:
10.14309/ctg.0000000000000014

Development and Validation of Confocal Endomicroscopy Diagnostic Criteria for Low-Grade Dysplasia in Barrett's Esophagus

Massimiliano di Pietro, MD[1], Helga Bertani, MD[2], Maria O'Donovan, MB, MD, FRCPath[3], Patricia Santos, MD[1], Hani Alastal, MPhil[1,4], Richard Phillips, BA (Cantab)[1], Jacobo Ortiz-Fernández-Sordo, MD[5], Marietta Iacucci, MD, PhD[6], Ines Modolell, MD[7], Luca Reggiani Bonetti, MD[8], Krish Ragunath, MD[5] and Lorenz Wernisch, PhD[9]

OBJECTIVES: **Low-grade dysplasia (LGD) in Barrett's esophagus (BE) is generally inconspicuous on conventional and magnified endoscopy. Probe-based confocal laser endomicroscopy (pCLE) provides insight into gastrointestinal mucosa at cellular resolution. We aimed to identify endomicroscopic features and develop pCLE diagnostic criteria for BE-related LGD.**

METHODS: **This was a retrospective study on pCLE videos generated in 2 prospective studies. In phase I, 2 investigators assessed 30 videos to identify LGD endomicroscopic features, which were then validated in an independent video set (n = 25). Criteria with average accuracy >80% and interobserver agreement $\kappa > 0.4$ were taken forward. In phase II, 6 endoscopists evaluated the criteria in an independent video set (n = 57). The area under receiver operating characteristic curve was constructed to find the best cutoff. Sensitivity, specificity, interobserver, and intraobserver agreements were calculated.**

RESULTS: **In phase I, 6 out of 8 criteria achieved the agreement and accuracy thresholds (i) dark nonround glands, (ii) irregular gland shape, (iii) lack of goblet cells, (iv) sharp cutoff of darkness, (v) variable cell size, and (vi) cellular stratification. The best cutoff for LGD diagnosis was 3 out of 6 positive criteria. In phase II, the diagnostic criteria had a sensitivity and specificity for LGD of 81.9% and 74.6%, respectively, with an area under receiver operating characteristic of 0.888. The interobserver agreement was substantial ($\kappa = 0.654$), and the mean intraobserver agreement was moderate ($\kappa = 0.590$).**

CONCLUSIONS: **We have generated and validated pCLE criteria for LGD in BE. Using these criteria, pCLE diagnosis of LGD is reproducible and has a substantial interobserver agreement.**

SUPPLEMENTARY MATERIAL accompanies this paper at http://links.lww.com/CTG/A23 and http://links.lww.com/CTG/A24

Clinical and Translational Gastroenterology 2019;00:e-00014. https://doi.org/10.14309/ctg.0000000000000014

INTRODUCTION

Barrett's esophagus (BE) is a premalignant condition to esophageal adenocarcinoma, with an estimated annual cancer risk of approximately 0.3% (1). Patients with BE are generally recommended endoscopic surveillance for early cancer detection (2,3). The cancer risk is significantly higher in dysplastic BE, with low-grade (LGD) and high-grade dysplasia (HGD) carrying a risk of progression of 0.5%–13% per year (4,5) and up to 2%–10% per year, respectively (6). A timely diagnosis of dysplasia is therefore essential for cancer prevention. Gastroenterology Societies recommend endoscopic ablation for both HGD and LGD (3,7), as there is randomized controlled evidence that this significantly reduces the rate of cancer progression (8).

The diagnosis of BE-related dysplasia, particularly LGD, is difficult as this is generally inconspicuous on the conventional and magnified endoscopic imaging. For this reason, it is recommended that multiple random biopsies be taken according to the Seattle protocol to allow sufficient tissue sampling for the

[1]MRC Cancer Unit, University of Cambridge, Cambridge, United Kingdom; [2]Digestive Endoscopy Unit, Baggiovara, Azienda Ospedaliero-Universitaria di Modena, Italy; [3]Department of Histopathology, Cambridge University Hospitals, Cambridge, United Kingdom; [4]Life Sciences, University of South Wales, Pontypridd, Wales; [5]NIHR Nottingham Digestive Diseases Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, Nottingham, United Kingdom; [6]Institute of Translational of Medicine and NIHR Biomedical Research Centre, University of Birmingham, United Kingdom; [7]Department of Gastroenterology, Cambridge University Hospitals, Cambridge, United Kingdom; [8]Department of Pathology, Policlinico di Modena, University of Modena and Reggio Emilia, Modena, Italy; [9]MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom. **Correspondence:** Massimiliano di Pietro, MD. E-mail: md460@mrc-cu.cam.ac.uk.

Received July 10, 2018; accepted December 14, 2018; published online April 11, 2019

histopathological diagnosis (2,3). The caveats of the random biopsy protocol are long procedure time, poor tolerance by the patient, limited adherence by the endoscopists, and high cost for the health care system (9). Thus, several endoscopic techniques have been investigated to improve targeting biopsies (10). However, to date none of the enhanced imaging techniques has replaced random biopsies in routine practice (2). In particular, there is no evidence that LGD can be diagnosed by enhanced magnified and nonmagnified endoscopic modalities.

Probe-based confocal laser endomicroscopy (pCLE) has the advantage to provide microscopic views at cellular resolution of the gastro-intestinal mucosa to allow real-time histological diagnosis (10). Although pCLE provides point imaging and is subject to the same limitation of sampling error as biopsy forceps, it provides the possibility to swipe across the mucosal plane to interrogate a larger mucosal surface than a standard biopsy. pCLE has been validated for a diagnosis of HGD in BE, with 6 pCLE criteria being described: saw-toothed epithelial surface, enlarged cells, pleomorphic cells, not equidistant glands, glands unequal in size and shape, and lack of goblet cells (11). These yielded a diagnostic accuracy of 81.5%, with a substantial interobserver agreement ($\kappa = 0.61$). In a cross-sectional study on 101 patients, the combination of pCLE and high-resolution endoscopy had a sensitivity and a specificity for HGD/intramucosal cancer (IMC) of 93.5 and 67.1%, respectively (12). We previously used these diagnostic criteria in a small cross-sectional cohort study (n = 56), where pCLE had a sensitivity and a specificity for any grade of dysplasia of 96% and 74%, respectively (13). However, in the per-location analysis the sensitivity was lower at 83%, suggesting that the diagnostic criteria need refinement to allow a more robust diagnosis of LGD. Thus, the aim of this study were (i) to identify endomicroscopic features of BE-related LGD and (ii) to develop and validate pCLE diagnostic criteria for LGD in BE.

METHODS

Study design

Good quality video sequences were retrieved from 2 video libraries of pCLE procedures performed previously within 2 ethically approved studies evaluating patients with BE with and without dysplasia. The Cambridge library derived from a prospective single center study, which recruited 55 nonconsecutive patients with BE between 2012 and 2014 (13). pCLE was performed on the endoscopic areas targeted by autofluorescence imaging marked by argon plasma coagulation, with targeted biopsies taken within the same area to ensure correspondence between pCLE location and histologic assessment. This library consisted of over 300 video sequences from 194 endoscopic locations, of which 24 had LGD histology, 21 had HGD, and 7 IMC. The rate of LGD visible on white light endoscopy in this library was 14%. The Modena library derived from a single-center study recruiting 100 consecutive patients with BE between 2011 and 2012, of which 50 received pCLE diagnosis (14). The correspondence between pCLE and targeted biopsy location was ensured by immediate tissue sampling sequential to pCLE analysis. In the pCLE group, 14 patients received an LGD diagnosis. None of the cases of LGD included in this library corresponded to the lesions visible on WLE.

Similarly to the work conducted for the development of HGD criteria (11), this study consisted of 2 phases. The first phase aimed to the generation of the diagnostic criteria for LGD and their initial external validation, while the second phase aimed to the definitive external validation of the criteria by a larger panel of investigators with different level of experience in pCLE.

The reference standard was considered the histological diagnosis on the targeted biopsies, which was scored prior to the result of the index test (pCLE). Dysplasia was graded according to the Vienna classification (15) and confirmed by at least 2 independent pathologists with extensive experience in the diagnosis of BE (M.O.D. and L.R.B.). In case of discordance between the 2 pathologists, the cases were excluded from the study.

All video sequences were selected and collated by an investigator, who completed the online pCLE training module, but had no previous experience in pCLE.

Generation and preliminary validation of the diagnostic criteria (phase I)

Thirty good quality videos (non-dysplastic BE (NDBE), n = 10; LGD, n = 10; and HGD, n = 10) from the Cambridge library were assessed in an unblinded manner by an endoscopist with previous experience in pCLE (M.d.P.) and a pathologist with extensive experience in BE diagnosis (M.O.D.) to identify recurrent pCLE features of LGD (Figure 1). We included HGD videos in this phase so that criteria, which are specific for LGD, could be identified. These 2 investigators independently assessed each of these criteria in a different set of good quality videos (NDBE, n = 15; LGD, n = 10), blinded to the histologic diagnosis. The criteria were scored in a binary fashion to assess for each of their sensitivity, specificity, and interobserver agreements. Only the criteria with accuracy (average of sensitivity and specificity) >80% and κ-value > 0.4 were taken forward. The candidate criteria were further evaluated on the same video set by a third investigator with previous experience in pCLE diagnosis (H.B.) to ensure reproducibility of the interpretation, before taking the criteria to the

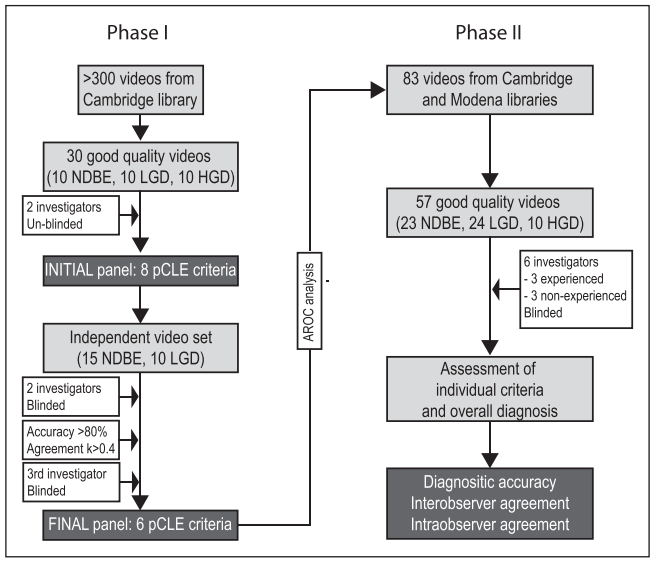

Figure 1. Schematic representation of study flowchart. In phase I (left), 2 investigators (one pathologist and one probe-based confocal laser endomicroscopy (pCLE)-experienced endoscopist) identified recurrent endomicroscopic features in low-grade dysplasia by blinded assessment of pCLE video sequences. Three investigators (one pathologist and 2 pCLE-experienced endoscopists) performed the initial validation of the criteria for the selection of the final diagnostic panel. In phase II (right), 6 endoscopists with different level of experience in pCLE performed the external validation on an independent video set (n = 57). The diagnostic accuracy and the interobserver agreement were calculated.

second phase. A sensitivity and specificity analysis for various cutoffs on the number of criteria for a positive call of LGD was performed to determine a useful range of cutoffs.

Validation of the diagnostic criteria (phase II)

For the purpose of the external validation, an independent set of 57 good and intermediate quality videos (NDBE, n = 23; LGD, n = 24; HGD, n = 10) were selected from the Modena and Cambridge libraries (Figure 1). The intermediate quality videos were allowed in this phase as long as they contained interpretable frames for optical diagnosis. Cambridge library videos used in phase II were different from those used in phase I. The videos were coded and collated in random order into a power point presentation. Six endoscopists, of which 3 had previous pCLE experience (at least 10 pCLE procedures for BE diagnosis) and 3 had no previous pCLE expertise, evaluated the diagnostic criteria obtained in phase I. Prior to the video examination, a training session was held with a detailed explanation of the pCLE features relative to the diagnostic criteria and joint assessment of 20 videos from phase I. In each video sequence, the 6 investigators, blinded to the histologic diagnosis, assessed each diagnostic criterion in a binary fashion (positive/negative) and made an overall diagnosis (dysplasia/no dysplasia), indicating the level of confidence (high/low). During the assessment, the investigators had the possibility to pause and rewind as it is done during live pCLE examination with the trackball. Although it was sufficient for a single frame to make a call of positive criterion, it was the discretion of the investigator to evaluate the significance of this and also based on the quality of the video and the presence of artifacts. We did not expect each case of LGD to have all 6 criteria positive as the videos were randomly selected. The assessors were not asked to make a differential diagnosis between LGD and HGD, as this was beyond the scope of this study. After the first 20 videos, a 30-minute interactive discussion was held to address the discrepancies on the diagnoses and compare them with the gold standard histopathology. Following this, the investigators assessed the remaining 37 videos.

Sample size and power analysis

There has been no previous study assessing the diagnostic accuracy of pCLE for LGD. Previously, a panel of pCLE criteria had a sensitivity of 75.6% for HGD. Our preliminary study suggested that the same diagnostic panel had a sensitivity in the per-location analysis of 83% (13). Assuming 80% as the true sensitivity for dysplasia of a panel of pCLE criteria, at least 18 dysplastic cases are required to formally show a sensitivity of at least 55% at a significance level of 0.05 and with a power of 0.7. To keep the data set balanced, we aimed for a similar number of controls.

Statistical analysis

Sensitivity, specificity, and accuracy were calculated as averages over all investigators. The lower bounds of the upper 95% credible intervals are provided for sensitivity and specificity. The credible interval for k out of n calls is obtained as the upper 95% quantile of the Beta ($k+1$, $n-k+1$) distribution (which is the posterior with a uniform Beta prior), where k is the average number of correct calls of all investigators and n is the number of samples with LGD for sensitivity and the number control samples for specificity. Intraobserver agreement was calculated using Cohen's κ statistic. Interobserver agreement among investigators was calculated using Fleiss' κ statistic. For the analyses assessing an agreement for minimum number of positive criteria for LGD diagnosis (Tables 2 and 4), a cutoff of n positive criteria was applied, regardless of what criteria were called positive. The differences between experienced and nonexperienced investigators in κ statistic were statistically assessed assuming approximate normality of the estimator for each group using its standard error. The differences between accuracies for these 2 groups were assessed with a t test on the logit-transformed values. Area under receiver operating characteristic (AROC) denotes the area under the convex hull of the receiver operating curve of sensitivity plotted over specificity for all cutoffs on the number of investigators making a call and is an additional measure alongside the average accuracy for the joint performance of all investigators. All analyses were performed in the R statistical environment.

RESULTS

Phase I

Unblinded assessment of dysplastic and nondysplastic cases (NDBE, n = 10; LGD, n = 10; and HGD, n = 10) by one pathologist and one endoscopist led to the identification of 8 recurrent confocal features in LGD cases. These were (I) dark nonround glands, (II) irregular gland shape, (III) lack of goblet

Table 1. Performance of the initial 8 diagnostic criteria in the blinded external validation (phase I) by 2 investigators (in brackets the lower bound of the upper 95% credible interval)

Criterion	Average sensitivity (%)	Average specificity (%)	Average accuracy (%)	Fleiss' kappa	AROC
Dark nonround glands	80 (53.0)	83.3 (61.9)	81.7	0.754	0.867
Irregular gland shape	85 (58.1)	90 (69.5)	87.5	0.667	0.93
Lack of goblet cells	75 (48.2)	86.7 (65.6)	80.8	0.576	0.88
Disrupted honeycomb pattern	30 (13.5)	90 (69.5)	60	−0.22	0.7
Variable degree of darkness of glandular epithelium with sharp cutoff	80 (53.0)	80 (58.3)	80	0.513	0.87
Variable size of cells	75 (48.2)	90 (69.5)	82.5	0.479	0.91
Pencil-shaped nuclei	70 (43.6)	63.3 (42.1)	66.7	0.44	0.713
Cellular stratification	85 (58.1)	86.7 (65.6)	85.8	0.754	0.9

AROC, area under the receiving operating characteristics curve.

Table 2. Performance of different cutoffs on 6 pCLE diagnostic criteria in the blinded validation (phase I) by 3 endoscopists (in brackets the lower bound of the upper 95% credible interval)

Cut-off	Average sensitivity (%)	Average specificity (%)	Average accuracy (%)	Fleiss' kappa	AROC
≥1 positive criterion	93.3 (67.5)	55.6 (35.2)	74.4	0.363	0.893
≥2 positive criteria	90.0 (63.6)	68.9 (47.3)	79.4	0.677	0.883
≥3 positive criteria	83.3 (56.4)	82.2 (60.7)	82.8	0.675	0.917
≥4 positive criteria	83.3 (56.4)	86.7 (65.6)	85	0.725	0.923
≥5 positive criteria	76.7 (49.7)	95.6 (76.5)	86.1	0.82	0.9
=6 positive criteria	63.3 (37.8)	95.6 (76.5)	79.4	0.537	0.9
Overall diagnosis	83.3 (56.4)	86.7 (65.6)	85	0.725	0.923

AROC, area under the receiving operating characteristics curve; pCLE, probe-based confocal laser endomicroscopy.

cells, (IV) disrupted honeycomb pattern, (V) variable degree of darkness with sharp cutoff, (VI) variable size of cells, (VII) pencil-shaped nuclei, and (VIII) cellular stratification. These 8 criteria were assessed in an independent set of 25 videos by the same 2 investigators blinded to the histologic diagnosis. The overall sensitivity, specificity, accuracy, and interobserver agreements and AROC for each criterion are shown in Table 1. Two of these criteria (disrupted honeycomb pattern and pencil-shaped nuclei) were discarded as they did not meet the minimum thresholds of 80% accuracy and $\kappa > 0.4$. The remaining 6 represented the final panel (Figure 2; see Figure 1, Supplementary Digital Content 2, http://links.lww.com/CTG/A24). An internal validation of these 6 criteria was performed by a second endoscopist experienced in pCLE. The performance of the criteria was similar, with only 2 criteria showing a slightly lower overall diagnostic accuracy (73.3% for variable degree of darkness with sharp cutoff and 74.6% for variable size of cells). Analyzing the results for all 3 investigators, we assessed the optimal cutoff for the minimum number of positive criteria to make a call of LGD as shown in Table 2. Apart from applying the pCLE criteria, the investigators also provided an "overall diagnosis" of LGD or no dysplasia, which is also shown in Table 2. The highest average accuracy (86.1%) and interobserver agreement (0.82) was found for a cutoff of 5 positive criteria, the highest AROC (0.923) for a cutoff of 4. However, even a cutoff of 3 criteria showed a high sensitivity and an AROC of 0.917, which was close to the optimum. In view of the need for a high sensitivity of the procedure, we opted for a final cutoff of 3 out of 6 criteria to be validated in phase II of this study.

Phase II

The 6 criteria identified in phase I were evaluated by 6 endoscopists (3 experienced and 3 non-experienced in pCLE), blinded to histologic diagnosis and any other clinical information, on a set of 57 videos (NDBE, n = 23; LGD, n = 24; and HGD, n = 10). The performance of each of the 6 criteria, the "overall diagnosis," and the diagnosis obtained applying a cutoff of ≥ 3 positive criteria are shown in Table 3. The accuracy of each criterion ranged from 69.2% to 77.6%, with an overall agreement ranging from poor to good (κ value between 0.273 and 0.667). The "overall diagnosis" and the cutoff of ≥3 positive criteria had accuracy of 79.1% and 78.3%, and AROC values of 0.921 and 0.888, respectively. Notably, the sensitivity slightly increased with the cutoff ≥3 positive criteria, while the overall accuracy was the same compared with "overall diagnosis"; that is, a cutoff procedure has the desirable effect of increasing sensitivity. It also increases the interobserver agreement, which is highest (0.654) with the cutoff of 3 positive criteria. There was a marginal improvement in the agreement levels after the interactive discussion, which was not statistically significant (κ value increased from 0.540 to 6.444, $P = 0.11$). The intraobserver agreement for the overall diagnosis showed a large degree of variation from fair to almost perfect across different investigators (Table 4). When we looked at the

Table 3. Performance of the selected 6 pCLE criteria for a diagnosis of LGD in the blinded validation (phase II, 23 NDBE and 24 LGD cases) by 6 endoscopists (the lower bound of the upper 95% credible interval is shown in brackets)

Criterion	Average sensitivity (%)	Average specificity (%)	Average accuracy (%)	Fleiss' kappa	AROC
Dark nonround glands	64.6 (47.6)	79 (61.9)	71.8	0.624	0.815
Irregular gland shape	79.9 (63.2)	75.4 (58)	77.6	0.667	0.865
Lack of goblet cells	63.9 (46.9)	76.1 (58.8)	70	0.422	0.83
Variable degree of darkness with sharp cut-off	68.1 (51)	71.7 (54.3)	69.9	0.327	0.857
Variable size of cells	65.3 (48.3)	73.2 (55.8)	69.2	0.273	0.864
Cellular stratification	63.9 (46.9)	76.8 (59.5)	70.4	0.349	0.863
Overall diagnosis	79.9 (63.2)	78.3 (61.1)	79.1	0.639	0.921
Cutoff ≥ 3 positive criteria	81.9 (65.5)	74.6 (57.3)	78.3	0.654	0.888

AROC, area under the receiving operating characteristics curve; LGD, low-grade dysplasia; pCLE, probe-based confocal laser endomicroscopy.

Table 4. Intraobserver agreement for the overall diagnosis and the individual diagnostic criteria

Overall diagnosis		Individual criteria		
Investigator	Cohen's κ	Criterion	Mean Cohen's κ	IQR
1	0.836	Dark nonround glands	0.616	0.25
2	0.587	Irregular gland shape	0.585	0.46
3	0.729	Lack of goblet cells	0.457	0.45
4	0.490	Variable darkness, sharp cutoff	0.320	0.32
5	0.550	Variable size of cells	0.293	0.54
6	0.349	Cellular stratification	0.346	0.44

For individual investigators, the intraobserver agreement for the overall diagnosis was assessed by Cohen's κ statistics. For individual criteria, the mean κ values are shown with the IQR.
IQR, interquartile range.

individual criteria, similarly to what observed for inter-rater analysis, the intraobserver agreement was better for the first 3 criteria and varied on average from fair to moderate (Table 4). In order to exclude that the new LGD criteria could miss a diagnosis of HGD, which could be detrimental for the utility of pCLE in the assessment of the overall dysplasia, we also included 10 HGD cases in the phase II validation. Reassuringly, the sensitivity for any grade of dysplasia slightly increased to 84.3% for the "overall diagnosis" and 84.8% for the cutoff of ≥3 positive criteria (see Table 1, Supplementary Digital Content 1, http://links.lww.com/CTG/A23). The sensitivity for HGD only was 95.0% for the "overall diagnosis" and 91.7% for the cutoff of ≥3 positive criteria (see Table 2, Supplementary Digital Content 1, http://links.lww.com/CTG/A23). These data indicate that the new LGD criteria can also accurately diagnose HGD.

Finally, we were interested in the effect of previous pCLE experience in performance (Table 5). The endoscopists experienced in pCLE performed slightly better than the nonexperienced ones on all statistics, even though this did not reach statistical significance ($P = 0.1$ for overall diagnosis and $P = 0.08$ for cutoff ≥3 positive criteria). Interestingly, the cutoff compared to the overall diagnosis increased sensitivity for experienced investigators (from 83.3% to 88.9%), but not for nonexperts (from 76.4% to 75%).

DISCUSSION

In this study, we applied a robust methodology to develop and validate novel pCLE diagnostic criteria for LGD in BE.

The American Society of Gastrointestinal Endoscopy program for the Preservation and Incorporation of Valuable Endoscopic Innovations (PIVI) has set the minimum thresholds of sensitivity, specificity, and negative predictive value (NPV) of novel imaging modalities for a diagnosis of HGD/IMC at 90%, 80%, and 98%, respectively. A recent meta-analysis showed that pCLE has sensitivity, NPV, and specificity for a diagnosis of HGD and IMC of 90.3%, 95.1%, and 77.3%, respectively (16). Hence, the minimum NPV and specificity PIVI thresholds of 98% and 80% were not met (17). Virtually, all the cases of IMC and approximately 2/3 of cases of HGD are associated to endoscopically visible lesions (18,19); hence, the role of pCLE to inform clinical management is limited in these cases. In contrast, LGD in BE is generally inconspicuous on white light endoscopy and even enhanced imaging techniques and represents only about 10% of visible neoplasia (20). Hence, applying the PIVI threshold to endoscopic diagnosis of LGD would be far too stringent. Nevertheless, LGD correlates with a significantly increased risk of cancer progression (5,21,22), indicating that an endoscopic diagnosis of LGD is essential and that availability of imaging techniques to aid this diagnosis is warranted. To date, there are no studies that have addressed the accuracy of enhanced imaging modalities specifically for a diagnosis of LGD.

LGD is associated with cellular and architectural features that are often difficult to interpret even by expert pathologists. As a result, interobserver agreement among pathologists is poor or moderate at best (22). In our study, the interobserver agreement for the 6 diagnostic criteria ranged in phase I from moderate to

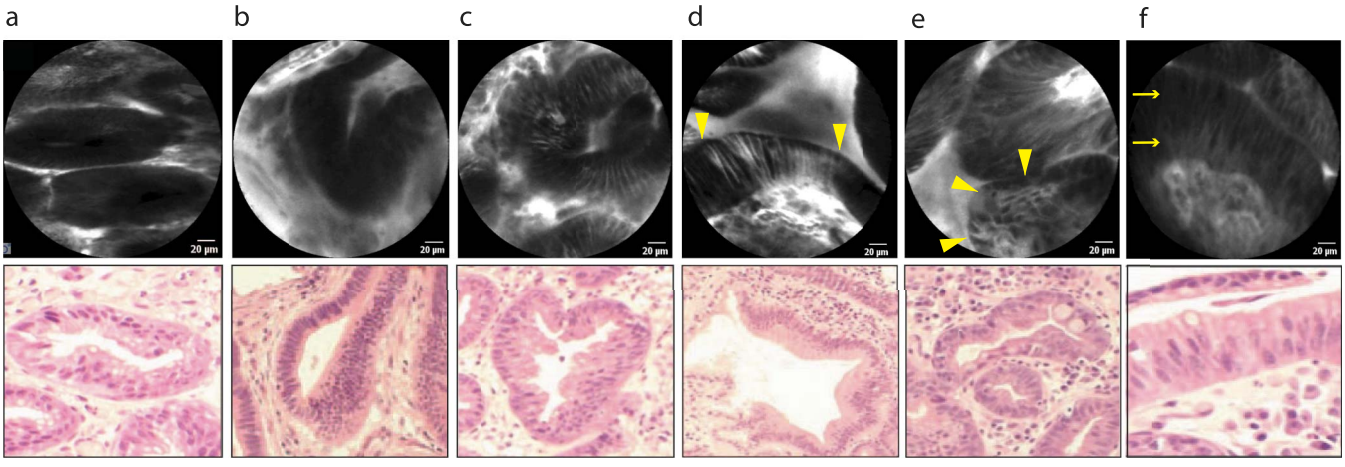

Figure 2. Final probe-based confocal laser endomicroscopy (pCLE) diagnostic panel, with pCLE view (top) and histological finding from the same endoscopic area (bottom). (**a**) dark nonround glands, (**b**) irregular gland shape, (**c**) lack of goblet cells, (**d**) variable degree of darkness with sharp cutoff (yellow arrowhead indicate transition points), (**e**) variable cell size (yellow arrowheads indicate the site of cellular pleomorphism), and (**f**) cellular stratification (yellow arrows indicate the 2 levels where cells are located). Although precise correspondence between glandular elements showed in the confocal and pathology image is difficult to be ascertained, similar features were observed on pCLE and histology.

Table 5. Performance of the panel of pCLE criteria for an LGD diagnosis using either a cutoff of 3 out 6 positive criteria (cutoff 3/6) or the overall diagnosis

Investigators/confidence	Average sensitivity (%)	Average specificity (%)	Average accuracy (%)	Fleiss' kappa	AROC	*P* value
Experienced—overall diagnosis	83.3 (67.0)	79.7 (62.6)	81.5	0.716	0.902	0.1
Nonexperienced—overall diagnosis	76.4 (59.5)	76.8 (59.5)	76.6	0.546	0.858	
Experienced—cut-off 3/6	88.9 (73.5)	75.4 (58.0)	82.1	0.71	0.905	0.08
Nonexperienced—cut-off 3/6	75.0 (58.0)	73.9 (56.5)	74.5	0.603	0.845	

Shown are results for the 3 experienced investigators, the 3 nonexperienced investigators (the lower bound of the upper 95% credible interval is shown in brackets).
Video set including 23 NDBE and 24 LGD cases.
AROC, area under the receiving operating characteristics curve; LGD, low-grade dysplasia; pCLE, probe-based confocal laser endomicroscopy.

substantial and in phase II from poor to substantial. However, it was very encouraging that when the criteria were used as a panel, the agreement among the endoscopists was substantial ($\kappa = 0.654$).

We found some overlapping pCLE features between HGD and LGD, such as the lack of goblet cells, the irregular gland structures, and the cellular pleomorphisms. This was not surprising given that the pathological cutoffs between LGD and HGD occur along a continuum with clear distinctions at either end of the spectrum but with somewhat overlapping features at the upper end of LGD, which is one of the reasons for the well-documented interobserver and intraobserver variability in the grading of dysplasia in BE. However, compared with HGD/IMC, the confocal features of LGD are subtler and the level of architectural and cellular distortion far less pronounced, which explains the lack of saw-toothed epithelial surface and non-equidistant glands in LGD cases. We believe that the dark glands in dysplastic cases reflect the high cellularity in dysplasia, which does not allow penetration of fluorescein in the intercellular spaces, and the variable degree of darkness reflects a sudden transition in the level of cellularity and stratification. Overall, it was encouraging that the pCLE criteria for LGD performed equally well to the validated pCLE criteria for HGD/IMC. In particular, compared to the HGD criteria, the LGD has a slightly better sensitivity (81.9% vs 75%), but lower specificity (74.6% vs 85.2%). The sensitivity increased when we included the cases of HGD, suggesting that the revised pCLE criteria can be used for a diagnosis of any grade of dysplasia, which is very useful in clinical practice.

This study has several strengths. We applied a strict methodology for the identification and the validation of the diagnostic criteria, with validation of these by larger panel of endoscopists. The good diagnostic accuracy achieved in phase II indicates that the diagnostic panel can be successfully applied even by nonexperienced operators. In addition, we found a clear correspondence between the recurrent confocal features in cases with LGD and the histologic appearances in the same cases. This indicates that it is very unlikely that the confocal criteria represent imaging artifacts. Finally, we used a robust pathological diagnosis of dysplasia with review of the cases by at least 2 pathologists and exclusion of cases where agreement was not achieved.

However, this study has also some limitations. The images were retrieved retrospectively, and 2 of the investigators were involved in the diagnostic procedures. Nevertheless, the original video libraries were historical and the retrieval of the videos was performed by investigators who did not take part in diagnostic procedures and validation. This ensured that the observers were blinded to the histologic diagnosis. Another limitation is the lack of endoscopic images corresponding to the confocal videos. This is a useful adjunct in clinical practice; however, it can also influence the pretest diagnosis. Furthermore, although all the efforts were made in the original trials to ensure correspondence between location of the video analysis and forceps biopsy, we cannot exclude that in some cases histology may not accurately represent the cellular structures shown in the videos. This is an inherent limitation of all endoscopy studies assessing magnified imaging modalities. Finally, we did not evaluate the differential diagnosis between LGD and HGD by pCLE. The primary aim of the study was to develop diagnostic criteria for LGD. Given the well-described difficulties in the differential diagnosis between LGD and HGD on conventional pathology and the fact that the clinical management of LGD and HGD is now very similar (3,7), we think that it is not possible nor clinically useful to distinguish different degrees of dysplasia by pCLE.

In summary, we have developed novel pCLE criteria for a diagnosis of LGD in BE and we propose that these should be used in future prospective studies aiming to endoscopic diagnosis of LGD.

CONFLICTS OF INTEREST

Guarantor of the article: Massimiliano di Pietro, MD.

Specific author contributions: study concept and design: M.d.P., H.B., M.O.D., and L.W. Acquisition of data: M.d.P., H.B., M.O.D., J.O.-F.-S., M.I., I.M., L.R.B., and K.R. Analysis and interpretation of data: M.d.P., M.O.D., and L.W. Statistical analysis: M.d.P., R.P., and L.W. Drafting of the article: M.d.P. and L.W. Technical and material support: P.S. and H.A. Final approval of the manuscript: All authors.

Financial support: The Addenbrookes Charitable Trust and the Kathy Shaw Memorial (Oesophageal Cancer) Fund funded the endoscopy equipment at Cambridge University Hospital. This study received infrastructure support from the Experimental Cancer Medicine Center and from the Cambridge Cancer Centre.

Potential competing interests: None to report.

ACKNOWLEDGMENTS

We thank Wladyslaw Januszewicz for the help with selecting pCLE video sequences; Irene Debiram, Bincy Alias, and Tara Nuckcheddy (MRC Cancer Unit, University of Cambridge) for their help with patient recruitment in Cambridge.

Study Highlights

WHAT IS KNOWN

- ✓ LGD in BE carries a significant cancer risk.
- ✓ Currently, there are no imaging modalities to aid this diagnosis.
- ✓ pCLE allows in vivo view of the gastrointestinal mucosa at cellular resolution.

WHAT IS NEW HERE

- ✓ We identified 6 pCLE features that are recurrently associated with LGD in BE, 3 of which have never been described before.
- ✓ As a panel, these diagnostic criteria led to the diagnosis of LGD in BE with 81.9% sensitivity and 74.6% specificity.
- ✓ The interobserver agreement for the overall pCLE diagnosis of LGD was substantial, indicating steep learning curve.

TRANSLATIONAL IMPACT

- ✓ The new pCLE criteria for diagnosis of LGD in BE can be used in future prospective studies.

REFERENCES

1. Desai TK, Krishnan K, Samala N, et al. The incidence of oesophageal adenocarcinoma in non-dysplastic Barrett's oesophagus: A meta-analysis. Gut. 2012;61(7):970–6.
2. Fitzgerald RC, di Pietro M, Ragunath K, et al. British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. Gut. 2014;63(1):7–42.
3. Shaheen NJ, Falk GW, Iyer PG, et al. ACG Clinical Guideline: Diagnosis and management of Barrett's esophagus. Am J Gastroenterol. 2016; 111(1):30–51; quiz 1.
4. Bhat S, Coleman HG, Yousef F, et al. Risk of malignant progression in Barrett's esophagus patients: Results from a large population-based study. J Natl Cancer Inst. 2011;103(13):1049–57.
5. Duits LC, Phoa KN, Curvers WL, et al. Barrett's oesophagus patients with low-grade dysplasia can be accurately risk-stratified after histological review by an expert pathology panel. Gut. 2015;64(5):700–6.
6. Rastogi A, Puli S, El-Serag HB, et al. Incidence of esophageal adenocarcinoma in patients with Barrett's esophagus and high-grade dysplasia: A meta-analysis. Gastrointest Endosc. 2008;67(3):394–8.
7. di Pietro M, Fitzgerald RC, BSG Barrett's Guidelines Working Group. Revised British Society of Gastroenterology recommendation on the diagnosis and management of Barrett's oesophagus with low-grade dysplasia. Gut. 2017;67(2):392–3.
8. Phoa KN, van Vilsteren FG, Weusten BL, et al. Radiofrequency ablation vs endoscopic surveillance for patients with Barrett esophagus and low-grade dysplasia: A randomized clinical trial. JAMA. 2014;311(12):1209–17.
9. Abrams JA, Kapel RC, Lindberg GM, et al. Adherence to biopsy guidelines for Barrett's esophagus surveillance in the community setting in the United States. Clin Gastroenterol Hepatol. 2009;7(7): 736–42.
10. Coda S, Thillainayagam AV. State of the art in advanced endoscopic imaging for the detection and evaluation of dysplasia and early cancer of the gastrointestinal tract. Clin Exp Gastroenterol. 2014;7: 133–50.
11. Gaddam S, Mathur SC, Singh M, et al. Novel probe-based confocal laser endomicroscopy criteria and interobserver agreement for the detection of dysplasia in Barrett's esophagus. Am J Gastroenterol. 2011;106(11): 1961–9.
12. Sharma P, Meining AR, Coron E, et al. Real-time increased detection of neoplastic tissue in Barrett's esophagus with probe-based confocal laser endomicroscopy: Final results of an international multicenter, prospective, randomized, controlled trial. Gastrointest Endosc. 2011; 74(3):465–72.
13. di Pietro M, Bird-Lieberman EL, Liu X, et al. Autofluorescence-directed confocal endomicroscopy in combination with a three-biomarker panel can inform management decisions in Barrett's esophagus. Am J Gastroenterol. 2015;110(11):1549–58.
14. Bertani H, Frazzoni M, Dabizzi E, et al. Improved detection of incident dysplasia by probe-based confocal laser endomicroscopy in a Barrett's esophagus surveillance program. Dig Dis Sci. 2013;58(1):188–93.
15. Dixon MF. Gastrointestinal epithelial neoplasia: Vienna revisited. Gut. 2002;51(1):130–1.
16. Committee AT, Thosani N, Abu Dayyeh BK, et al. ASGE Technology Committee systematic review and meta-analysis assessing the ASGE Preservation and Incorporation of Valuable Endoscopic Innovations thresholds for adopting real-time imaging-assisted endoscopic targeted biopsy during endoscopic surveillance of Barrett's esophagus. Gastrointest Endosc. 2016;83(4):684–98.e7.
17. Sharma P, Savides TJ, Canto MI, et al. The American Society for Gastrointestinal Endoscopy PIVI (Preservation and Incorporation of Valuable Endoscopic iInnovations) on imaging in Barrett's esophagus. Gastrointest Endosc. 2012;76(2):252–4.
18. Haidry RJ, Butt MA, Dunn J, et al. Radiofrequency ablation for early oesophageal squamous neoplasia: Outcomes form United Kingdom registry. World J Gastroenterol. 2013;19(36):6011–9.
19. Phoa KN, Pouw RE, Bisschops R, et al. Multimodality endoscopic eradication for neoplastic Barrett oesophagus: Results of an European multicentre study (EURO-II). Gut. 2016;65(4):555–62.
20. Wani S, Abrams J, Edmundowicz SA, et al. Endoscopic mucosal resection results in change of histologic diagnosis in Barrett's esophagus patients with visible and flat neoplasia: A multicenter cohort study. Dig Dis Sci. 2013;58(6):1703–9.
21. Hvid-Jensen F, Pedersen L, Drewes AM, et al. Incidence of adenocarcinoma among patients with Barrett's esophagus. N Engl J Med. 2011;365(15):1375–83.
22. Wani S, Falk GW, Post J, et al. Risk factors for progression of low-grade dysplasia in patients with Barrett's esophagus. Gastroenterology. 2011; 141(4):1179–86.