The learning curve, interobserver, and intraobserver agreement of endoscopic confocal laser endomicroscopy in the assessment of mucosal barrier defects

Jeff Chang, BSc(Med), MBBS, FRACP,[1,2] Matthew Ip,[1,2] Michael Yang,[1,2] Brendon Wong,[1,2] Theresa Power, BMedSci, BMBS, FRCPA,[3,4] Lisa Lin, BSc(Med), MBBS, FRCPA,[3,4] Wei Xuan, MSc, MAppStat, PhD,[5] Tri Giang Phan, MBBS, FRACP, FRCPA, PhD,[2,6] Rupert W. Leong, MBBS, FRACP, MD, AGAF[1,2,6,7]

Sydney, New South Wales, Australia

Background and Aims: Confocal laser endomicroscopy can dynamically assess intestinal mucosal barrier defects and increased intestinal permeability (IP). These are functional features that do not have corresponding appearance on histopathology. As such, previous pathology training may not be beneficial in learning these dynamic features. This study aims to evaluate the diagnostic accuracy, learning curve, inter- and intraobserver agreement for identifying features of increased IP in experienced and inexperienced analysts and pathologists.

Methods: A total of 180 endoscopic confocal laser endomicroscopy (Pentax EC-3870FK; Pentax, Tokyo, Japan) images of the terminal ileum, subdivided into 6 sets of 30 were evaluated by 6 experienced analysts, 13 inexperienced analysts, and 2 pathologists, after a 30-minute teaching session. Cell-junction enhancement, fluorescein leak, and cell dropout were used to represent increased IP and were either present or absent in each image. For each image, the diagnostic accuracy, confidence, and quality were assessed.

Results: Diagnostic accuracy was significantly higher for experienced analysts compared with inexperienced analysts from the first set (96.7% vs 83.1%, $P < .001$) to the third set (95% vs 89.7, $P = .127$). No differences in accuracy were noted between inexperienced analysts and pathologists. Confidence (odds ratio, 8.71; 95% confidence interval, 5.58-13.57) and good image quality (odds ratio, 1.58; 95% confidence interval, 1.22-2.03) were associated with improved interpretation. Interobserver agreement κ values were high and improved with experience (experienced analysts, 0.83; inexperienced analysts, 0.73; and pathologists, 0.62). Intraobserver agreement was >0.86 for experienced observers.

Conclusion: Features representative of increased IP can be rapidly learned with high inter- and intraobserver agreement. Confidence and image quality were significant predictors of accurate interpretation. Previous pathology training did not have an effect on learning. (Gastrointest Endosc 2016;83:785-91.)

In the normal bowel, mucosal barrier integrity is maintained by a monolayer of epithelial cells held together by tight junctions.[1,2] Under homeostatic conditions, epithelial cells are shed from the tips of the villi in a controlled process, involving reorganization of the tight junctions to maintain integrity of the epithelial barrier.[3] However,

Abbreviations: CDO, cell dropout; CI, confidence interval; CJE, cell-junction enhancement; CLE, confocal laser endomicroscopy; CLS, confocal leak score; eCLE, endoscopic confocal laser endomicroscopy; FL, fluorescein leak; IBD, inflammatory bowel disease; IP, intestinal permeability; OR, odds ratio.

DISCLOSURE: *All authors disclosed no financial relationships relevant to this publication.*

0016-5107/$36.00
http://dx.doi.org/10.1016/j.gie.2015.08.045

Received January 27, 2015. Accepted August 26, 2015.

Current affiliations: Gastroenterology and Liver Services, Bankstown Hospital, Sydney South West Local Health District (1), Faculty of Medicine, University of New South Wales (2), Department of Histopathology, Douglass Hanly Moir Pathology (3), Faculty of Medicine, The University of Notre Dame (4), Ingham Institute of Applied Medical Research (5), Immunology Division, The Garvan Institute of Medical Research (6), Australian School of Advanced Medicine, Macquarie University (7), Sydney, New South Wales, Australia.

Reprint requests: Rupert W. Leong, MBBS, FRACP, MD, AGAF, Gastroenterology and Liver Services, Bankstown-Lidcombe and Concord Hospital, Concord Hospital, Level 1 West, Hospital Rd., Concord, NSW 2137 Australia.

If you would like to chat with an author of this article, you may contact Dr Leong at rupert.leong@sswahs.nsw.gov.au.

when pathological cell shedding increases and exceeds the rate of repair, large gaps lead to loss of epithelial integrity and increased intestinal permeability (IP). The resultant exposure of luminal antigens to host immune cells leads to immune activation and inflammatory cell recruitment, which further drive cell shedding.[4] Previous studies have suggested that impaired epithelial barrier function is present in both inflammatory bowel disease (IBD) and irritable bowel syndrome, with possible pathoetiological implications of disease.[2,5-7] Indirect assays for measuring IP have quantified urinary excretion of chromium ethylenediamine tetraacetic acid or saccharide probes such as sucrose, lactulose, and mannitol after ingestion.[8] These methods have illustrated that both Crohn's disease and ulcerative colitis have higher IP.[9] Although simple and noninvasive, these tests have the limitations of low sensitivity and specificity, without the capacity to discriminate between inflamed and noninflamed tissues. Furthermore, inaccuracies could occur due to confounding factors of differential rates of bacterial degradation, intestinal transit time, and renal excretion.[10]

Endoscope-based confocal laser endomicroscopy (eCLE) integrates a confocal laser microscope within the tip of a standard flexible endoscope, allowing simultaneous macroscopic and microscopic assessment of GI mucosa. It uses a blue laser emitter to scan focal planes of target mucosa at 1000-fold magnification at varying depths of up to 250 μm to allow real-time in vivo histology assessments.[11,12] When this is used to obtain virtual biopsy specimens, CLE provides images equivalent to histological assessment in Barrett's esophagus, colorectal cancer, and celiac disease.[13-15] Learning characteristics and observer reliability of CLE in assessing varying diseases have been previously reported.[20-24] Although the learning characteristics of static images are established, no previous study has investigated the learning curve and inter- and intraobserver reliability for assessing eCLE features of a dynamic process such as increased IP. Also, the extent to which an analyst's experience and confidence may influence diagnostic accuracy and reliability remains unclear. Pathologists who are trained in the recognition of static histopathology may not have any advantage in the learning of the dynamic features of increased IP. Dynamic features representative of increased IP include cell-junction enhancement (CJE), fluorescein leak (FL) and cell dropout (CDO), but they cannot be visualized on conventional histopathology.[16]

The primary aim of this study was to evaluate the diagnostic accuracy, learning curve, inter- and intraobserver agreement for correctly identifying eCLE images of increased IP between analysts with and without previous training and pathologists. The secondary aim was to examine the impact of analysts' experience, image quality, and analysts' confidence as predictors of diagnostic accuracy.

METHODS

Data collection

Previously, 101 patients (80 patients with IBD, 21 control patients) prospectively recruited from the IBD Service of Bankstown-Lidcombe Hospital, Sydney, Australia, underwent eCLE (Pentax EC-3870FK; Pentax, Tokyo, Japan) with 5 mL of intravenous 10% fluorescein sodium given in increments. Diagnosis of IBD was made according to clinical, radiological, and histological findings. Patients were invited to participate in the study if they were older than 18 years of age, able to provide informed consent, and required a colonoscopy for clinical indications. Predetermined exclusion criteria were pregnancy, kidney disease, or known allergies to fluorescein. All eCLE images were taken from macroscopically normal tissue from the terminal ileum confirmed with paired histology yielding a total of 45,882 deidentified images. These were stored in a digital database at Bankstown-Lidcombe Hospital, of which 180 images (57 control, 123 increased IP) were selected by the reference eCLE analyst (R.W.L.) for this study. All of the 57 control images were derived from the 21 control patients, with the 123 increased IP features derived from 36 of the 80 patients with IBD, who represented those with the most significant barrier dysfunction. No more than 5 images were selected from an individual patient. Each image was assessed for the presence or absence of features of increased IP, as well as image quality by the same reference eCLE analyst who did not participate in the subsequent evaluation. A "good quality" image had at least one-third of usable material without blurred movement artifacts. The study was approved by Human Research Ethics Committee of Sydney South Western Area Health Service (reference number 14/327).

Reference standard

The 3 eCLE features of increased IP described previously are CJE, a buildup of fluorescein between 2 epithelial cells representing impaired tight-junction proteins before breakage of the final basal tight junction releasing the fluorescein into the lumen; FL, a fluorescein plume entering the lumen from between 2 enterocytes representing loss of apposition between 2 adjacent cells; and CDO, shedding of an apoptotic enterocyte into the luminal space (Fig. 1).[17] All features were initially derived through bench-top intravital 2-photon confocal laser microscopy of explant murine small intestines bathed in luminal fluorescein. Mice were given intravenous Hoechst 33258 to stain epithelial cellular nuclei as well as either intraperitoneal phosphate buffer solution as controls or 5 μg of tumor necrosis factor-α to induce intestinal mucosal cell shedding. Under confocal laser microscopy imaging (Carl Zeiss 7MP; Carl Zeiss, Oberkochen, Germany and Chameleon Vision II Ti:Sa laser; Coherent Scientific, Hilton, South Australia,

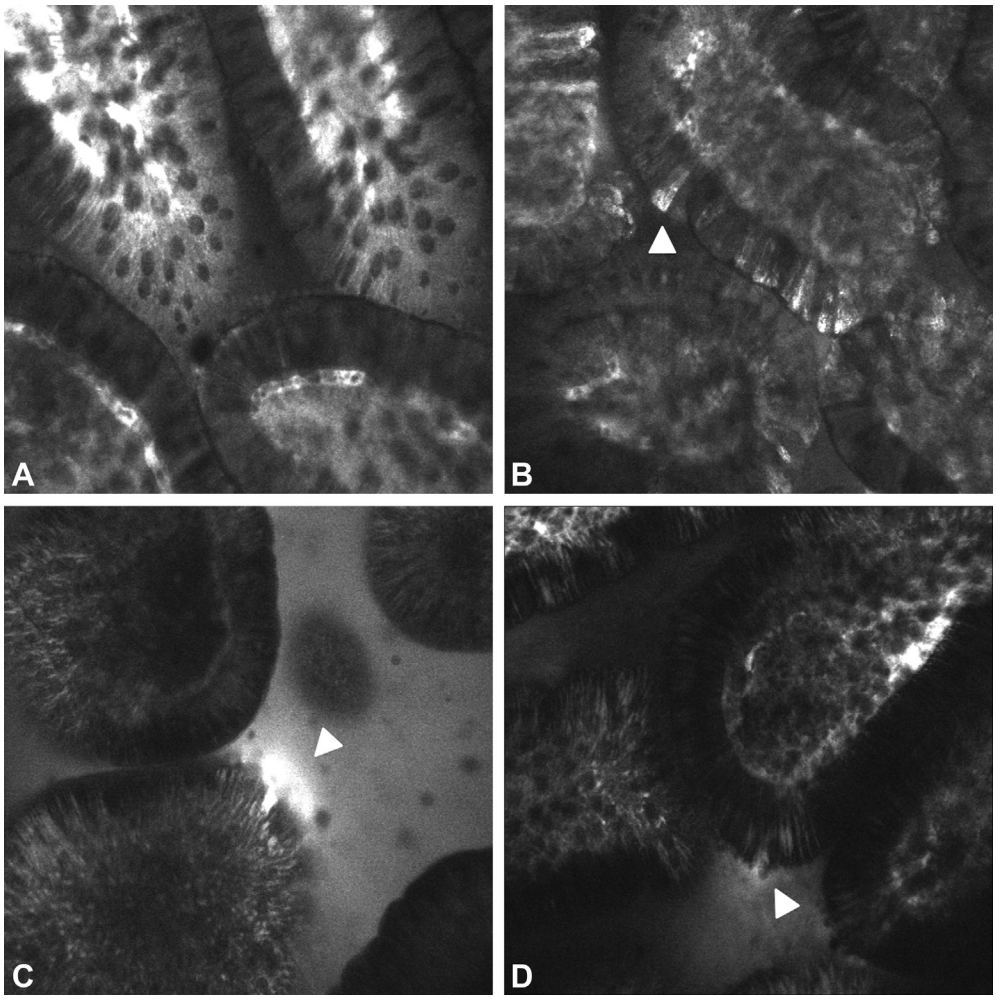

Figure 1. Endoscopic confocal laser endomicroscopic features of increased intestinal permeability. Control (**A**), cell-junction enhancement (**B**), fluorescein leak (**C**), cell dropout (**D**).

Australia), epithelial shedding and FL were demonstrated with quantification of cell shedding by the image analysis program IMARIS. Pilot comparative assessments of the mucosal barrier in human terminal ileum by eCLE in controls and patients with IBD were used to prospectively identify equivalent features through a focus group meeting of eCLE experts. Consensus was reached to not include features with poor interobserver agreement (for example, gaps and microerosions), leaving the 3 aforementioned features.[16,18] To quantify the severity of barrier dysfunction, a composite numerical score, the confocal leak score (CLS), was developed, which is derived from the total number of images demonstrating CJE, FL, and CDO proportional to total images used represented as a percentage. This provided for the first time a continuous grade of severity rather than a dichotomous variable, with higher values demonstrating a greater degree of barrier dysfunction. This linear scale allowed for the recognition of physiological levels of mucosal cell shedding, which cannot be quantified by using the dichotomous Watson grading system.

Training and image evaluation

Before the evaluation, a brief didactic 30-minute training session was held to cover the criteria for image interpretation. This included a learning set of 12 images (3 images each of CJE, FL, and CDO, and 3 control images) for which immediate feedback was provided. Training images were not repeated in the final evaluation that would follow.

For the assessment of the learning characteristics, the 180 images selected were divided into 6 sets of 30 images, with each set containing equal number of images assessing CJE, CDO, and FL. Analysts were defined as experienced based on their minimum experience of 100 eCLE examinations, inexperienced identified as those with no previous eCLE image interpretation, and pathologists as board-certified practicing consultants experienced in GI histopathology without previous eCLE image interpretation experience. All analysts were blinded to both patient diagnosis and the number of images evaluated in each set for specific eCLE features. For each image, analysts indicated the presence or absence of a specific feature (CJE, FL, or

CDO) and the confidence of their response as either confident or a guess. Only 1 feature was present in each image. At the conclusion of each set, a review of the completed set was conducted followed by a 15-minute rest break before embarking on the next set. The same method of assessment was conducted for experienced analysts after 6 months to determine the intraobserver agreement.

Statistical analysis

The accuracy, sensitivity, and specificity of each analyst class were calculated as the mean of each analyst's score within the group. Two-sample *t* tests and 1-way analysis of variance were used to compare mean diagnostic accuracies between groups. The learning curve for inexperienced analysts and pathologists was illustrated by an unpaired *t* test and defined as the point at which no significant difference can be detected in the accuracy of interpreting IP between inexperienced analysts and experienced analysts. Generalized estimating equations were used to calculate the odds ratio (OR) and 95% confidence interval (CI) of the impact of image quality, confidence, and experience on diagnostic accuracy. The interobserver agreement for each of the 3 study groups was calculated by the Fleiss and Cohen κ statistics. The intraobserver reliability was calculated for experienced observers 6 months after the index assessment by Cohen's κ statistics. Strengths of agreement were acknowledged as a guess (κ = 0), slight (κ = 0.01-0.20), fair (κ = 0.21-0.40), moderate (κ = 0.41-0.60), substantial (κ = 0.61-0.80), almost perfect (κ = 0.81-0.99), and perfect (κ = 1.00) agreement. *P* values <.05 were considered statistically significant. Statistical analyses were performed by using IBM SPSS Statistics 20.0 (SPSS Inc, Chicago, Ill).

RESULTS

Subjects

Twenty-one analysts were prospectively recruited, comprising 6 eCLE experienced analysts, 13 inexperienced analysts, and 2 pathologists.

Diagnostic accuracy and learning curve

The overall sensitivity, specificity, and accuracy for the 3 groups are outlined in Table 1. The experienced and inexperienced analysts and the pathologists had overall mean accuracies of 95.9%, 87.8%, and 85.6%, respectively ($P < .001$), in interpreting eCLE features of increased IP. Significant differences were identified between experienced and inexperienced analysts (95.9% vs 87.8%, $P < .001$), and between experienced analysts and pathologists (95.9% vs 85.6%, $P < .001$). The odds of correctly interpreting an image was higher for experienced analysts compared with inexperienced analysts (OR, 2.67; 95% CI, 1.36-5.22; $P = .004$) and pathologists (OR, 3.06; 95% CI, 1.61-5.83; $P = .001$). No statistical differences were noted between pathologists compared with inexperienced analysts (OR, 0.87; 95% CI, 0.66-1.15; $P = .325$).

TABLE 1. Accuracy, sensitivity, and specificity of experienced/inexperienced analysts and pathologists

Image no.	Analyst class	Sensitivity, %	Specificity, %	Accuracy, %
1-30	Experienced	95.3	97.9	96.7
	Inexperienced	80.9	85.0	83.1
	Pathologists	78.6	87.5	83.3
31-60	Experienced	88.1	97.9	93.3
	Inexperienced	73.6	85.6	80.0
	Pathologists	57.1	84.4	71.7
61-90	Experienced	97.1	93.6	95.6
	Inexperienced	88.2	89.3	88.7
	Pathologists	94.1	84.6	90.0
91-120	Experienced	96.3	94.4	95.0
	Inexperienced	90.7	89.3	89.7
	Pathologists	100.0	85.7	90.0
121-150	Experienced	99.0	98.7	98.9
	Inexperienced	90.5	99.4	94.4
	Pathologists	91.2	84.6	88.3
151-180	Experienced	93.1	98.1	96.1
	Inexperienced	87.8	92.7	90.8
	Pathologists	88.0	91.4	90.0
Total	Experienced	95.4	96.4	95.9
	Inexperienced	85.2	90.0	87.8
	Pathologists	84.4	86.5	85.6

The learning curve was assessed by comparing mean accuracies between the groups at the end of each set (Fig. 2). Differences in accuracy were noted between experienced and inexperienced analysts in the first (96.7% vs 83.1%, $P < .001$), second (93.3% vs 80%, $P < .001$), and third sets (95.6% vs 88.7%, $P < .001$). This difference, however, was no longer statistically significant by the end of set 4 ($P = .127$) and remained so for both sets 5 and 6 (Supplementary Table 1). No differences in accuracy could be demonstrated between pathologists and inexperienced analysts either at the end of each individual set or overall (87.8% vs 85.6%, $P = .309$).

Impact of image quality and analyst confidence on diagnostic accuracy

Analysis by generalized estimating equations demonstrated that both the image quality and confidence with which analysts interpreted the images were positive predictors of accurate identification of features of IP (OR, 1.58; 95% CI, 1.22-2.03; $P < .001$ and OR, 8.71; 95% CI, 5.58-13.57; $P < .001$, respectively) after adjusting for the effects of experience (Table 2). When assessing individual

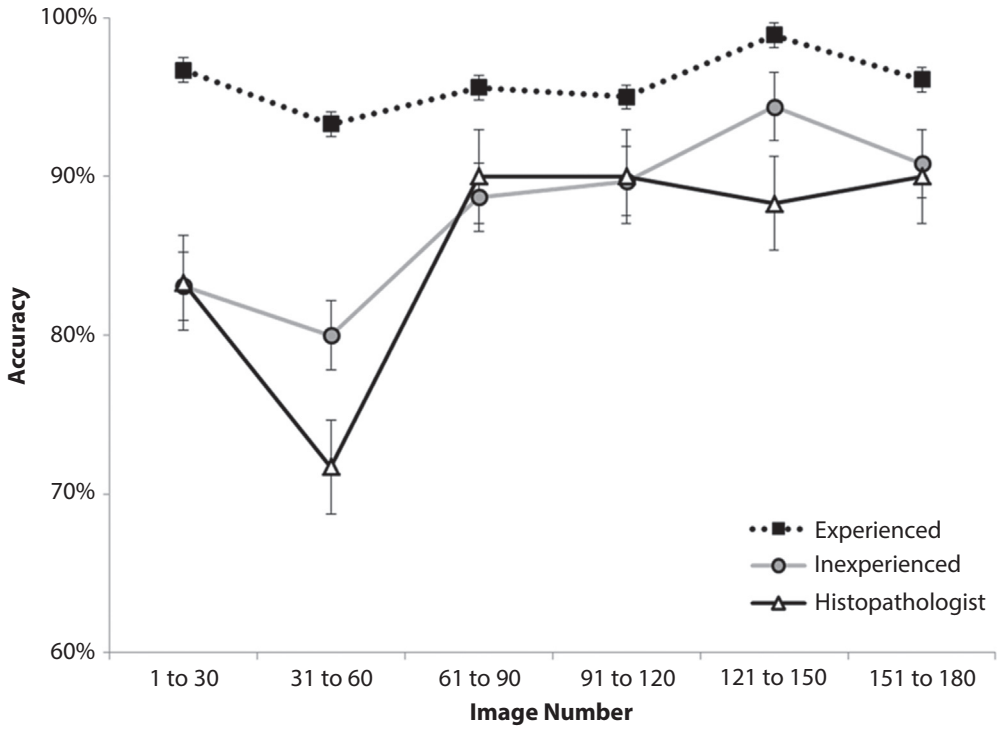

Figure 2. Learning curve as represented by diagnostic accuracy (percentage and 95% confidence interval) in experienced analysis (*black squares*), inexperienced analysts (*gray circles*), and pathologists (*open triangles*) at the end of each set.

TABLE 2. Generalized estimating equation: impact of image quality, confidence, and experience on diagnostic accuracy

Factors		Odds ratio (95% CI)	*P* value
Image quality	Normal	1.00	—
	Good	1.58 (1.22-2.03)	<.001
Response	Guess	1.00	—
	Confident	8.71 (5.58-13.57)	<.001
Level of experience	Inexperienced	1.00	—
	Pathologists	0.87 (0.66-1.15)	.325
	Experienced	2.67 (1.36-5.22)	.004

CI, Confidence interval.

groups, the impact of between-analyst confidence on their accuracy was significant for all groups, but more evident for inexperienced analysts (OR, 10.76; 95% CI, 6.12-18.89; $P < .001$), with a diagnostic accuracy of 94.8% achieved in questions answered confidently, as well as by pathologists (OR, 5.36; 95% CI, 2.86-10.05; $P < .001$) with an accuracy of 91.2% (Table 3).

Interobserver agreement

The overall interobserver agreement was substantial in all groups, with the highest agreement in the experienced group ($\kappa = 0.83$) compared with the inexperienced analysts and pathologists ($\kappa = .72$ and $\kappa = 0.62$, respectively). Interobserver variabilities for each individual feature in each group are shown in Table 4. When assessed independently, FL appeared to have the best interobserver agreement compared with CJE and CDO in all 3 groups. However, all 3 features are an integral aspect of the CLS measurement in assessing intestinal permeability given that on principal component analysis (data not shown), the contribution to CLS for each feature was independently variable and could not excluded. Improvement of agreement was demonstrated in both inexperienced analysts ($\kappa = 0.64$ to 0.79) and pathologists ($\kappa = .46$ to 0.66) from the first set to the final set.

Intraobserver agreement of experienced analysts

Paired analyses at 6 months identified substantial intraobserver agreement for all 6 experienced confocal laser endomicroscopy (CLE) analysts, with an overall mean κ value of 0.94 ($\pm$ 0.047 SD). Furthermore, consistently high intraobserver agreement κ values were achieved for each image set for each analyst (Table 5).

DISCUSSION

IP as a result of barrier defects is not shown by histopathology because ex vivo specimens taken by forceps are devitalized and cannot demonstrate physiological function. eCLE may represent an in vivo, reliable, and accurate method for identifying increased IP, which has been implicated in irritable bowel syndrome. Recently, eCLE was also shown to be effective in visualizing epithelial gaps that occur

TABLE 3. Generalized estimating equation: impact of image quality and confidence within individual groups

Factors		Inexperienced, odds ratio (95% CI)	Experienced, odds ratio (95% CI)	Pathologist, odds ratio (95% CI)
Image quality	Normal	1.00	1.00	1.00
	Good	1.52 (1.08-2.12)*	1.93 (1.28-2.90)*	1.79 (1.27-2.52)*
Response	Guess	1.00	1.00	1.00
	Confident	10.76 (6.12-18.89)*	4.58 (1.59-13.18)*	5.36 (2.86-10.05)*

CI, Confidence interval.
*$P < .05$.

TABLE 4. Interobserver agreement (κ) for different analyst class and individual features

Analyst class	Overall interobserver agreement (κ)	FL interobserver agreement (κ)	CJE interobserver agreement (κ)	CDO interobserver agreement (κ)
Experienced	0.83	0.92	0.85	0.80
Inexperienced	0.72	0.92	0.59	0.52
Pathologist	0.62	0.77	0.57	0.54

FL, Fluorescein leak; *CJE*, cell-junction enhancement; *CDO*, cell dropout.

TABLE 5. Intraobserver agreement (κ) for experienced analysts

Experienced analysts	Intraobserver agreement (κ)
1	0.99
2	0.92
3	0.96
4	0.86
5	0.97
6	0.96

in response to food antigens in real time in patients with irritable bowel syndrome who had suspected food intolerance.[19] Increased IP on eCLE has also been implicated in IBD.[16] However, that study only validated increased IP to be either present or absent and did not demonstrate physiological levels of leak that occur in non-IBD controls, did not demonstrate reversibility or long- term prognosis, and was not scrutinized for learning characteristics.[16] The learning curve, accuracy, and reliability of interpreting eCLE images of increased IP (CJE, FL, and CDO) have never been previously established. This is the first study to illustrate and compare the learning curve of inexperienced analysts and pathologists with experienced analysts in recognising increased IP on eCLE. We also demonstrated that previous pathology training was not advantageous in acquiring eCLE interpretation skills.

There is a short learning curve for interpreting eCLE images demonstrating features of increased IP with the loss of statistical difference between mean diagnostic accuracy between expert and inexperienced analysts after the teaching module and exposure to 90 images. Previous pathology training did not improve learning, indicating that interpretation of dynamic features on eCLE to be sufficiently different from conventional light microscopy. Due to these differences, previous pathology training may, in fact, interfere with the interpretation of IP. These results were also supported by another study that showed medical students untrained in eCLE image interpretation achieved higher κ scores than pathologists.[20]

Image quality and analyst confidence increased diagnostic accuracy, with analysts being 1.5 times as likely to correctly identify these features when image quality was good compared with normal quality and almost 9 times more likely to answer correctly when confident compared with a guessed response. This is keeping with previous studies that identified confidence to be correlated with higher diagnostic accuracy in the interpretation of esophageal and colorectal CLE images to their corresponding pathologies.[21-24]

The overall interobserver κ scores were higher in experienced analysts than inexperienced analysts and pathologists, with the latter 2 groups improving from the first to last set of images. FL had the best interobserver agreement in all 3 groups, but all 3 features are fundamental in the assessment of CLS as a measure of IP. Reproducibility with nearly perfect intraobserver κ scores of 0.86 to 0.99 was demonstrated. These results support use of eCLE as a reliable tool to assess epithelial barrier integrity by measuring increased IP.

Limitations need to be acknowledged. Unlike other diseases in which eCLE images can be compared with histology, features of increased IP possess no criterion standard equivalent for comparison. Although the use of eCLE to assess IP has previously been demonstrated in several studies, no previous studies have assessed the learning curve or interobserver variability.[25,26] In our study, a senior experienced eCLE analyst (R.W.L.), who was otherwise not involved in the analytical part of the study, was used as reference standard. The learning curve was still demonstrated by using this reference standard, and the high

accuracy of other experienced CLE analysts was demonstrated. Another limitation is the postprocedural analysis of eCLE images rather than a real-time assessment. However, the focus of our study was on the interpretation of eCLE features of increased IP and the learning curve, which was performed blinded to the endoscopic images in a controlled environment.

In conclusion, our study illustrates that the eCLE features of CJE, FL, and CDO can be easily learned with a short learning curve to achieve high levels of diagnostic accuracy. Accuracy increased with higher observer confidence, and interpretation was reliable both in the short and longer term. In addition, pathology training does not increase the learning curve or accuracy.

REFERENCES

1. Salim SY, Soderholm JD. Importance of disrupted intestinal barrier in inflammatory bowel diseases. Inflamm Bowel Dis 2011;17:362-81.
2. Turner JR. Intestinal mucosal barrier function in health and disease. Nat Rev Immunol 2009;9:799-809.
3. Watson AJ, Duckworth CA, Guan Y, et al. Mechanisms of epithelial cell shedding in the Mammalian intestine and maintenance of barrier function. Ann N Y Acad Sci 2009;1165:135-42.
4. Watson AJ, Hughes KR. TNF-alpha-induced intestinal epithelial cell shedding: implications for intestinal barrier function. Ann N Y Acad Sci 2012;1258:1-8.
5. McGuckin MA, Eri R, Simms LA, et al. Intestinal barrier dysfunction in inflammatory bowel diseases. Inflamm Bowel Dis 2009;15:100-13.
6. Irvine EJ, Marshall JK. Increased intestinal permeability precedes the onset of Crohn's disease in a subject with familial risk. Gastroenterology 2000;119:1740-4.
7. Paramsothy S, Leong RW. Endoscopy: fluorescein contrast in confocal laser endomicroscopy. Nat Rev Gastroenterol Hepatol 2010;7:366-8.
8. Arslan G, Atasever T, Cindoruk M, et al. 51CrEDTA colonic permeability and therapy response in patients with ulcerative colitis. Nucl Med Commun 2001;22:997-1001.
9. Arrieta MC, Bistritz L, Meddings JB. Alterations in intestinal permeability. Gut 2006;55:1512-20.
10. DeMeo MT, Mutlu EA, Keshavarzian A, et al. Intestinal permeation and gastrointestinal disease. J Clin Gastroenterol 2002;34:385-96.
11. Neumann H, Kiesslich R, Wallace MB, et al. Confocal laser endomicroscopy: technical advances and clinical applications. Gastroenterology 2010;139:388-92.
12. De Palma GD. Confocal laser endomicroscopy in the "in vivo" histological diagnosis of the gastrointestinal tract. World J Gastroenterol 2009;15:5770-5.
13. Nguyen NQ, Leong RW. Current application of confocal endomicroscopy in gastrointestinal disorders. J Gastroenterol Hepatol 2008;23:1483-91.
14. Othman MO, Wallace MB. Confocal laser endomicroscopy: is it prime time? J Clin Gastroenterol 2011;45:205-6.
15. Leong RW, Nguyen NQ, Meredith CG, et al. In vivo confocal endomicroscopy in the diagnosis and evaluation of celiac disease. Gastroenterology 2008;135:1870-6.
16. Kiesslich R, Duckworth CA, Moussata D, et al. Local barrier dysfunction identified by confocal laser endomicroscopy predicts relapse in inflammatory bowel disease. Gut 2012;61:1146-53.
17. Leong R, Wong B, Chen J, et al. Intestinal mucosal leakage is detected using in vivo confocal endomicroscopy in macroscopically-normal Crohn's disease and ulcerative colitis [abstract]. J Gastroenterol Hepatol 2012;27(Suppl 4):AB109.
18. Leong RW, Arshi M, Chang J, et al. Acute tumour necrosis factor administration induces transient epithelial cell shedding, mucosal leak and neutrophil recruitment to the gut in a mouse model. J Gastroenterol Hepatol 2014;29(Suppl 2):11.
19. Fritscher-Ravens A, Schuppan D, Ellrichmann M, et al. Confocal endomicroscopy shows food-associated changes in the intestinal mucosa of patients with irritable bowel syndrome. Gastroenterology 2014;147:1012-20.
20. Hoffman A, Rey JW, Mueller L, et al. Analysis of interobserver variability for endomicroscopy of the gastrointestinal tract. Dig Liver Dis 2014;46:140-5.
21. Buchner AM, Gomez V, Heckman MG, et al. The learning curve of in vivo probe-based confocal laser endomicroscopy for prediction of colorectal neoplasia. Gastrointest Endosc 2011;73:556-60.
22. Kuiper T, Kiesslich R, Ponsioen C, et al. The learning curve, accuracy, and interobserver agreement of endoscope-based confocal laser endomicroscopy for the differentiation of colorectal lesions. Gastrointest Endosc 2012;75:1211-7.
23. Wallace MB, Sharma P, Lightdale C, et al. Preliminary accuracy and interobserver agreement for the detection of intraepithelial neoplasia in Barrett's esophagus with probe-based confocal laser endomicroscopy. Gastrointest Endosc 2010;72:19-24.
24. Liu J, Li M, Li Z, et al. Learning curve and interobserver agreement of confocal laser endomicroscopy for detecting precancerous or early-stage esophageal squamous cancer. PLoS One 2014 4;9:e99089.
25. Kiesslich R, Goetz M, Angus EM, et al. Identification of epithelial gaps in human small and large intestine by confocal endomicroscopy. Gastroenterology 2007;133:1769-78.
26. Turcotte JF, Kao D, Mah SJ, et al. Breaks in the wall: increased gaps in the intestinal epithelium of irritable bowel syndrome patients identified by confocal laser endomicroscopy (with videos). Gastrointest Endosc 2013;77:624-30.

SUPPLEMENTARY TABLE 1. Comparison of diagnostic accuracy between experienced, inexperienced and pathologist analysts

Groups	Image no.	Accuracy, % Group 1	Group 2	*P* value
Experienced analysts (group 1) versus inexperienced analysts (group 2)	1-30	96.7	83.1	<.001
	31-60	93.3	80.0	<.001
	61-90	95.6	88.7	<.001
	91-120	95.0	89.7	.127
	121-150	98.9	94.4	.065
	151-180	96.1	90.8	.073
	Total	95.9	87.8	<.001
Experienced analysts (group 1) versus pathologists (group 2)	1-30	96.7	83.3	<.001
	31-60	93.3	71.7	<.001
	61-90	95.6	90.0	.003
	91-120	95.0	90.0	.205
	121-150	98.9	88.3	.001
	151-180	96.1	91.4	.011
	Total	95.9	85.6	<.001
Inexperienced analysts (group 1) versus pathologists (group 2)	1-30	83.1	83.3	.553
	31-60	80.0	71.7	.232
	61-90	88.7	90.0	.508
	91-120	89.7	90.0	1.000
	121-150	94.4	88.3	.066
	151-180	90.8	91.4	.309
	Total	87.8	85.6	.128