Title

INFLUENCE OF REVIEWERS' CLINICAL BACKGROUNDS ON INTERPRETATION OF CONFOCAL LASER ENDOMICROSCOPY FINDINGS

Short title

FACTORS INFLUENCING INTERPRETATION OF CLE

Authors

Masakuni Kobayashi[1,2]

Helmut Neumann[3]

Shoryoku Hino[4]

Michael Vieth[5]

Seiichiro Abe[6]

Yousuke Nakai[7]

Kiyokazu Nakajima[8]

Ralf Kiesslich[9]

Shinichi Hirooka[10]

Kazuki Sumiyama[1]

1. Department of Endoscopy, The Jikei University School of Medicine, Tokyo, Japan
2. Division of Gastroenterology and Hepatology, Department of Internal Medicine, The Jikei University School of Medicine, Tokyo, Japan
3. Department of Medicine, Interdisciplinary Endoscopy, University of Erlangen-Nuremberg, Erlangen, Germany
4. Department of Neuropsychiatry, Ishikawa Prefectural Takamatsu Hospital, Ishikawa, Japan
5. Institute for Pathology, Klinikum Bayreuth, Bayreuth, Germany
6. Endoscopy Division, National Cancer Center, Tokyo, Japan
7. Department of Gastroenterology, The University of Tokyo, Tokyo, Japan
8. Department of Surgery, Osaka University, Osaka, Japan
9. Helios HSK Clinic Wiesbaden, II. Department of internal medicine,

Wiesbaden, Germany

10. Department of Pathology, The Jikei University School of Medicine, Tokyo, Japan

Grant support

This work was supported by JSPS KAKENHI Grant Number 25860565.

Abbreviations

Abbreviations used in this paper: CLE, confocal laser endomicroscopy; NPV, negative predictive value; pCLE, probe-based confocal laser endomicroscopy ; PPV, positive predictive value; WLE, white-light endoscopy

Correspondence

Kazuki Sumiyama, M.D., Ph.D., Professor,

Department of Endoscopy, The Jikei University School of Medicine,

3-25-8 Nishi Shinbashi, Minato-ku, Tokyo 105-8461, Japan

E-mail: kaz_sum@jikei.ac.jp

Telephone number: 03-3433-1111 (ex 3181)

Fax number: 03-3459-4524

Conflicts of interest

Masakuni Kobayashi has no conflict to disclose.

Helmut Neumann has received scientific support from Olympus, Pentax Medical, Fujifilm, EndoChoice, Abbvie, Siemens, SpectraScience and Smart Medical. He has also given paid lectures for Pentax Medical, EndoChoice, Abbvie, SpectraScience, Aptalis, Mauna Kea Technologies, Eisai, Recordati, Takeda, Astra Zeneca and Falk, and is a consultant for Pentax Medical, Fujifilm, Smart Medical, EndoChoice, Fraunhofer and SpectraScience.

Shoryoku Hino has no conflict to disclose.

Michael Vieth has given paid lectures for Pentax Medical and Olympus, and unpaid lectures for Mauna Kea Technologies.

Seiichiro Abe has no conflict to disclose.

Yousuke Nakai has no conflict to disclose.

Kiyokazu Nakajima has no conflict to disclose.

Ralf Kiesslich received a research grant from Pentax Europe (for confocal laser endomicroscopy research).

Shinichi Hirooka has no conflict to disclose.

Kazuki Sumiyama has no conflict to disclose.

Authors' contributions

Masakuni Kobayashi carried out the study, performed the statistical analysis, and drafted the manuscript. Kazuki Sumiyama conceived the study, participated in its design and coordination, and helped to draft the manuscript. Helmut Neumann participated in the study and helped to draft the manuscript. Shoryoku Hino performed the statistical analysis. Michael Vieth and Shinichi Hirooka participated in the study and evaluated the pathological findings. Seiichiro Abe, Yousuke Nakai, Kiyokazu Nakajima, and Ralf Kiesslich participated in the study. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We are deeply grateful to the reviewers who participated in this study for their significant contributions, which enabled this international multi-center study, and to the patients who participated.

Abstract

Background and study aims: Substantial discrepancies in endoscopic strategy for gastric cancer exist between Western and Eastern countries owing to clinico-epidemiological diversity, including differences in the prevalence of gastric cancer. This international multicenter study involved German and Japanese institutions and aimed to evaluate the influence of reviewers' clinical backgrounds on interpreting probe-based confocal laser endomicroscopy (pCLE) findings for diagnosis of superficial gastric lesions.

Patients and methods: Thirty-nine reviewers answered questionnaires about their clinical background and then reviewed 30 sets of white-light endoscopy (WLE) and pCLE video clips via an online questionnaire. For each set of clips, reviewers were asked to classify lesions as neoplastic or non-neoplastic. Results of video reviews were compared with the final histopathological diagnosis for each lesion. The accuracy of diagnosis based on WLE + pCLE was compared with that based on WLE alone for each aspect of clinical background.

Results: The overall accuracy of diagnosis based on WLE + pCLE was higher

than that based on WLE alone (73.93% vs 65.64%, *P* = .0002). Outcomes of expert gastroenterologists were better than those of pathologists (*P* = .038 for WLE, *P* = .002 for WLE + pCLE) and outcomes of reviewers at Japanese institutions were better than those of reviewers at German institutions (*P* = .001 for WLE, *P* < 0.001 for WLE + pCLE).

Conclusions: Reviewers from Japanese institutions and expert gastroenterologists performed well in the pCLE interpretation. Substantial experience in conventional endoscopy is important for interpreting pCLE images for the diagnosis of gastric cancer.

INTRODUCTION

Confocal laser endomicroscopy (CLE) imaging has been shown to have equal diagnostic value to histopathological examination.[1-7] In a series of studies, it has been reported that CLE systems can help reduce the requirement for biopsies or help obtain biopsy specimens more efficiently in situations of low diagnostic yield,[8,9] and could be used to identify tumor differentiation and stages.[10-12] However, an image obtained with a CLE system is a horizontally sliced image, which differs from traditional histopathological assessment. In addition, a pathology-based diagnosis is established using a fixed sample, whereas a CLE-based diagnosis is uses moving images. Peter et al reported that there are discrepancies in the interpretation of probe-based CLE (pCLE) images pertaining to digestive diseases between endoscopists and pathologists.[13] We hypothesized that the interpretation of CLE images is substantially influenced by reviewers' clinical backgrounds, including specialty and endoscopy experience.

Currently, CLE systems are widely used for surveillance of Barrett's esophagus and ulcerative colitis in Europe and the United States.[1,2,9,14-21] However, few

reports are available on CLE-based diagnosis of gastric neoplasia from these countries.[10-12,22,23] Substantial discrepancies in endoscopic strategy for gastric cancer exist between Western and Eastern countries owing to diversity in clinico-epidemiolgical situations, including differences in the prevalence of gastric cancer.[24-26]

The aims of this study were to assess the influence of reviewers' clinical backgrounds on pCLE-based differential diagnosis of superficial gastric lesions (neoplastic and non-neoplastic) by collaborating with Western (German) and Eastern (Japanese) institutions. Although, there are two clinically available CLE systems — embedded CLE and pCLE — only pCLE is commercially available in Japan, so it was used in this study.

pCLE is usually performed using white-light endoscopy (WLE) in clinical practice. We therefore compared the accuracy of WLE alone with that of WLE plus pCLE.

MATERIALS AND METHODS

Patients

We included 30 consecutive patients (with a total of 45 gastric lesions) who underwent pCLE for gastric lesions since the pCLE system was introduced at our hospital (Jikei University Hospital, Tokyo, Japan). Between February 2013 and January 2014, WLE and fluorescein-assisted pCLE recordings of superficial gastric lesions were prospectively collected from the patients who met the following inclusion criteria: aged between 18 and 75 years, and able to give written informed consent. Patients with advanced malignant disease, an allergy to the fluorescent contrast agent fluorescein sodium, coagulopathy, a bleeding disorder, and/or severe liver or renal failure were excluded.

CLE system

We used a pCLE system (CellVizio, Mauna Kea Technologies, Paris, France) and a GastroFlex-UHD probe (Mauna Kea Technologies, Paris, France) with a 2.5 mm external diameter. This probe adapts to the accessory channel of an upper

endoscope and has a field view of 240 μm. pCLE images were obtained at a rate of 12 frames per second.

Endoscopic image acquisition

Patients drank a preparation of dimethicone (Gascon, Kissei Pharmaceutical Co, Ltd, Nagano, Japan) before examination. Endoscopy was performed under conscious sedation using mefenamic acid (Opistan, Tanabe Pharmaceutical Co, Tokyo, Japan), midazolam (Dormicum; Astellas Pharma Inc, Tokyo, Japan) or flunitrazepam (Rohypnol; Chugai Pharmaceutical Co, Ltd, Tokyo, Japan). At first, all patients underwent routine examination with magnifying endoscopy (GIF-Q260Z, Olympus Medical Systems Co, Tokyo, Japan). After carefully cleansing the surface of the lesion with gentle water lavage and removing as much mucous from the surface as possible, image acquisition was initiated by standard WLE observation of targeted areas. Targeted areas were observed by WLE for at least 30 seconds. Then, pCLE images were obtained from the center of targeted areas with intravenous administration of 5 mL of 10% fluorescein

sodium (Fluorescite Intravenous Injection 500 mg, Alcon Japan Ltd, Tokyo, Japan). Detailed pCLE observation of targeted areas was also performed for at least 30 seconds. All studied sites were biopsied immediately after observation. Biopsy specimens were taken from the center of the lesions or from the part of the mucosa that had scratch marks created by CLE probe contact during observation. All endoscopic procedures were recorded to a hard-disk drive and were performed by two expert gastroenterologists (KS and MK).

A set of 30-second demonstrative WLE and pCLE video clips of lesions from all 30 patients were edited by an expert gastroenterologist (MK) - the research coordinator of this study. Since we thought that the quality of pCLE images obtained by Japanese endoscopists who did not have high-volume experience in pCLE might not be high enough for accurate clinical interpretation, two gastroenterologists (KS and HN), one from Japan and one from Germany, who were blinded to the clinical information (including the histopathological diagnosis of lesions) selected 30 sets of high-quality video clips obtained from 30 lesions. Selections were solely based on image quality, and only sets with high-quality

WLE and high-quality pCLE images were chosen.

All clips were de-identified, and the research coordinator created a list of the clips that including the histopathological diagnosis for each, and uploaded the clips onto a web-based system. The research coordinator was the only person who was allowed to access to this system.

Review of video clips

All clips selected for use in the study were uploaded to an online questionnaire system (Survey Monkey, Palo Alto, Calif, USA), and reviewers registered to obtain a username and password that enabled them to access the questionnaire. The directors of each institution recruited volunteers to be reviewers in the study. All of those who volunteered to be reviewers participated in this study. After completing a web-based diagnostic tutorial on pCLE image interpretation (http://cellvizio.net/self-training, about 20 min duration), reviewers answered a questionnaire on their backgrounds, which included questions on country of practice, specialty, endoscopy experience, and pathology training (Table 1).

Subsequently, the reviewers watched the sets of selected clips via the online questionnaire system (in which the sequence of the sets of clips was randomly assigned for each reviewer) and classified each lesion as neoplastic or non-neoplastic. The reviewers always began with interpretation of WLE findings alone, because pCLE is normally performed after WLE in clinical practice. Revision of the initial diagnosis (ie, diagnosis based on interpreting WLE findings alone) was not allowed once the final diagnosis was made for each lesion (ie, diagnosis based on interpreting WLE + pCLE findings). In the reviewing process, the time allowed for interpreting the video clips was not limited. Reviewers needed to click a “Next” button to move to the next video clip, but there was no button that enabled them to go back to the previous video clip. A flow chart of the protocol is shown in Figure 1.

pCLE findings

pCLE findings of a completely disorganized epithelium, fluorescein leakage, or an abrupt change to “black cells” were defined as a neoplastic lesion (Figure 2);

these criteria are included in the Miami classification.[27]

Histopathological examination

All biopsy specimens and excised tissue samples were stained with hematoxylin and eosin. Microscopic images of all specimens were saved as JPEG files by the research coordinator and sent to two expert gastrointestinal pathologists (SH and MV) for histopathological diagnosis. In cases where there was a discrepancy, the histopathological diagnosis was discussed by e-mail. The pathologists were blinded to the endoscopic diagnosis based on the WLE and pCLE video clips. The histopathological diagnosis was reported according to the Vienna criteria for neoplasia[28]. If the superficial lesion was excised by endoscopic submucosal dissection or gastrectomy, the final histopathological diagnosis was based on the excised tissue.

Outcomes measures

The primary outcome of the study was the diagnostic accuracy of using WLE

findings alone and WLE + pCLE findings. Secondary outcomes included the overall accuracy of differential diagnosis by WLE and WLE + pCLE among the following reviewer groups: reviewers with CLE experience versus accuracy of reviewers with no CLE experience; German reviewers versus Japanese reviewers, gastroenterologists versus pathologists; and gastroenterologists with pathological training versus those without pathological training.

Statistical analysis

Data obtained from video clip reviews were analyzed using STATA 13 (Stata Corp LP, College Station, Tex, USA). A descriptive statistical method was used to analyze diagnostic accuracy. Overall accuracy, sensitivity, specificity, positive predictive values (PPV), negative predictive values (NPV), and positive and negative likelihood ratios for WLE and WLE + pCLE were calculated using the histopathological diagnosis (neoplastic or non-neoplastic) as the gold standard. The differences between WLE and WLE + pCLE were tested using the McNemar test for overall accuracy, sensitivity and specificity, and using the $\chi 2$ test for PPV

and NPV. *P* values < .05 were considered statistically significant. Inter-observer agreement was calculated using kappa statistics. To determine the factors associated with correct diagnosis, multilevel logistic regression was performed. The outcome variable was a concordance with the final histopathological diagnosis (neoplastic or non-neoplastic). Reviewer-level explanatory variables included country, specialty, with or without CLE experience, and with or without pathology training. Since each reviewer contributed multiple observations, the mixed model included reviewer as a random effect. We calculated odds ratios to assess the effect of each explanatory variable on likelihood a correct diagnosis.

Ethical considerations

The study protocol was approved by the ethics committee of the Jikei University School of Medicine (number 25-174 [7309]) and performed in accordance with the Declaration of Helsinki. Written informed consent was obtained from all patients included in the study before the procedure. All the authors had access to the study data and reviewed and approved the final manuscript. The study was

registered in the University Hospital Medical Information Network Clinical Trials Registry (UMIN-CTR; www.umin.ac.jp/ctr/index.htm; number UMIN 000013437).

RESULTS

From February 2013 to January 2014, 30 patients with 45 lesions were enrolled in this study, from whom 30 sets of WLE and pCLE video clips of superficial lesions in the stomach — 18 neoplastic and 12 non-neoplastic — were selected for inclusion in the online questionnaire (a total of 1170 reviewer–clip combinations). The characteristics of the lesions are summarized in Table 2. The average size of selected lesions was 11.7 mm. A total of 39 reviewers participated: seven from three German institutions and 32 from four Japanese institutions. Six reviewers were pathologists and 33 were gastroenterologists (Table 1). Thirty-two reviewers had no experience with CLE, while seven had CLE experience. Fourteen expert gastroenterologists with experience based on more than 1,000 gastrointestinal endoscopy cases had gone through pathology training, while 17 expert gastroenterologists had not trained in pathology.

The sensitivity, specificity, PPV, NPV and positive and negative likelihood ratio findings for WLE and WLE + pCLE are shown in Tables 3 and 4, respectively. The overall accuracy of the differential diagnosis for WLE and WLE + pCLE was 65.64% and 73.93%, respectively (Table 5). The accuracy of WLE + pCLE was generally higher than the accuracy of WLE alone, regardless of reviewer background. There were significant differences between the accuracies of WLE and WLE + pCLE when we compared gastroenterologists with pathologists (66.87% vs 51.32% for WLE [*P* = .038], 75.66% vs 64.44% [*P* = .002] for WLE + pCLE, respectively) and when we compared German reviewers with Japanese reviewers (55.71% vs 67.81% [*P* = .001] for WLE and 64.9% vs 73.21% [*P* < 0.001] for WLE + pCLE, respectively). However, there were no significant differences in diagnostic accuracy when we compared reviewers who had CLE experience with those who did not have CLE experience (64.29% vs 65.94% [*P* = .648] for WLE and 75.71% vs 73.54% [*P* = .516] for WLE + pCLE, respectively) and when we compared expert endoscopists who had pathology training with those who had not trained in pathology (67.38% vs 67.64% [*P* = .931] for WLE

and 73.33% vs 77.25% [*P* = .167] for WLE + pCLE, respectively). The diagnostic accuracies of three expert endoscopists with both pathology training and CLE experience for WLE and WLE + pCLE were 65.56% and 70.00%, respectively. These accuracies were not significantly different (*P = .316*). The diagnostic accuracies of WLE and WLE + pCLE for non-neoplastic lesions were significantly higher than those for neoplastic lesions (75.64% vs 58.97% [*P* = < .01] for WLE and 79.27% vs 70.37% [*P* = < .01] for WLE + pCLE, respectively).

Levels of inter-observer agreement between reviewers in each clinical background groups are shown in Table 6. Inter-observer agreement was generally higher for WLE + pCLE than for WLE.

In the multilevel logistic regression analysis, German institution and CLE experience were variables independently affecting diagnostic accuracy by WLE + pCLE (Table 7). The interpretation of pCLE was also identified as a variable independently influencing the overall diagnostic accuracy in addition to German institution and CLE experience (Table 8). The pCLE interpretation changed the initial WLE diagnosis in 24.02% (281/1170) of reviewer–clip combinations and led

to a correct answer in 67.26% (189/281) of those reviewer–clip combinations.

DISCUSSION

We investigated the influence of reviewers' clinical backgrounds on differential diagnostic accuracy using pCLE for superficial gastric lesions in Western and Eastern institutions. We also compared the overall accuracy of diagnoses based on WLE and WLE + pCLE. The overall accuracy for WLE + pCLE was higher than that for WLE, regardless of the reviewers' clinical backgrounds. Both direct comparisons and multivariate analysis clearly demonstrated that combining WLE examination with pCLE was more effective than WLE examination alone for the differential diagnosis of superficial gastric lesions. Also, correct answers were given by reviewers in two-thirds of instances in which they changed their answers based on pCLE findings.

Peter et al speculated that while endoscopists have training in real-time imaging, pathologists have an inherent advantage in terms of knowledge of pathological cellular differentiation.[13] They evaluated differences in interpretation of pCLE

findings between endoscopists and pathologists and found poor agreement between them — sensitivity, specificity and accuracy for diagnosis using pCLE were higher for endoscopists than for pathologists. In our study, the accuracy of diagnosis using WLE + pCLE was higher for gastroenterologists than for pathologists. Although we also evaluated the influence of pathology training on pCLE interpretation by gastroenterologists, such training had no effect on pCLE interpretation. The reason for this could be that image orientation in pCLE is differs from that in pathology images. Pathology images are fixed sample images and pathology findings rely on nuclear and structural atypia; in contrast, pCLE relies on moving images and pCLE-based diagnosis is only made from structural atypia. Also, due to a lack of blinding between pCLE and WLE images, the sequence of the interpretation might have had a negative effect on accuracy, especially for pathologists. However, pCLE interpretation always follows WLE interpretation in clinical settings.

In our study, reviewers' CLE experience did not influence pCLE image interpretation. Buchner et al examined the learning curve for correctly identifying

benign and neoplastic colorectal lesions with pCLE. They demonstrated that the prediction of colorectal neoplasia could be learned rapidly by a wide range of endoscopists with a 2-hour training session.[29] In our study, all reviewers received a 20 min tutorial before watching the WLE and WLE + pCLE video clips. The training might have contributed sufficiently to their understanding of pCLE interpretation to negate the diagnostic advantage of clinical CLE experience before this study.

We compared the accuracy of diagnosing neoplastic lesions across countries (Japan vs Germany). There were significant differences between Japanese and German reviewers in terms of accuracy of diagnoses based on WLE and WLE + CLE. Because of the higher incidence of gastric cancer in Japan,[30] Japanese gastroenterologists have a better chance of correctly establishing an endoscopic diagnosis of gastric cancer with meticulous gastric preparation using dimethicone in Japan. Consistent with this, we found that Japanese reviewers achieved better results than German reviewers in the WLE interpretation. Also, accuracy for pCLE was significantly higher for Japanese reviewers than German reviewers.

From the results of the multilevel logistic regression analysis, group-level variables of German institution and CLE experience negatively affected the diagnostic accuracy for WLE and WLE + pCLE. Meanwhile, the group-level variable of WLE + pCLE positively affected the diagnostic accuracy for WLE and WLE + pCLE. Although pCLE interpretation was identified as an independent variable in the multivariate analysis, and improved diagnostic accuracy, there were no significant differences between accuracy of diagnoses based on WLE and WLE + pCLE in the German group. The results indicate that correct WLE interpretation and adequate training in WLE is most important for achieving highly accurate diagnoses even with the use of pCLE for gastric cancer. Moreover, our results indicate the importance of such training and strengthen the rationale for standardized training under the guidance of endoscopy societies.

Bok et al compared the accuracy of endoscopic forceps biopsy and pCLE for diagnosis of superficial gastric neoplasia before endoscopic treatment. They reported that overall agreement with the final diagnosis based on pathology analysis was significantly higher for pCLE diagnosis than for biopsy diagnosis.

The overall accuracy of pCLE-based diagnosis of adenocarcinoma was 90.7%.[31] Zhang et al evaluated the sensitivity, specificity, and accuracy of gastric cancer diagnosis with embedded CLE, and found that the diagnostic accuracy was 97.1%.[11] In our study, the accuracy of WLE + pCLE-based diagnosis was 73.93% — significantly lower than in other studies. However, our study was designed to assess the influence of reviewers' clinical backgrounds; it was not designed to evaluate the accuracy of using pCLE for gastric cancer diagnosis. Also, interpreting video clips differs from real-time diagnosis in clinical settings. The reviewers in our study had a wide variety of clinical backgrounds, and some were pCLE novices. Consequently, it might not have been possible for the overall accuracy of WLE + pCLE-based diagnosis in our study to be comparable with that reported in other studies. However, pCLE-based diagnosis for gastric lesions seems to be influenced by WLE-based diagnosis, and the reliability of pCLE may not be the same for other indications.

In the analysis of inter-observer agreement between reviewers, levels of agreement for WLE + pCLE were generally higher than those for WLE, even

though results of the multilevel logistic regression analysis indicated that CLE experience negatively affected the diagnostic accuracy for WLE + pCLE. We surmised that gastric pCLE interpretation might require a specific knowledge base and classification to achieve more accurate diagnoses. Alternatively, the pCLE clips reviewed in this study might have shown unusual findings, which could have mislead the reviewers and thereby led to incorrect answers.

Our study had several limitations. First, a potential bias exists in the WLE and pCLE video clip selection, since video clips were short-listed by one expert endoscopist. Second, a sample size calculation was not used. We recruited volunteers for image reviewing and all those who volunteered to take part were included in the analysis – this approach was used to minimize selection bias in terms of reviewer selection. Third, considering the overall low accuracy, sensitivity and specificity noted in this study, a controlled study design with larger samples would enable stronger conclusions, especially in terms of elucidating the clinical value of pCLE. However, this would have been technically and financially difficult because pCLE was not approved by the Japanese government for clinical

use or reimbursement at the time of the study (it has since been approved for clinical use). Moreover, we thought that the study would be more informative if it was conducted when most Japanese endoscopists were naive to pCLE as this could help determine whether regular endoscopic experience has an influence on pCLE interpretation. The CLE system is not widely available in Japan, so most of the Japanese reviewers did not have experience in CLE. It would be useful to conduct further research on the details and quality of the pathology training system, and to confidentially assess the reviewers' answers. Regular embedded pathology training for gastroenterologists might not be enough to improve diagnostic accuracy.

Although we showed that pCLE enables a more accurate diagnosis of superficial gastric lesions, it has some disadvantages. First, CLE requires intravenous fluorescent reagents, and involves a longer procedure. Second, pCLE observational activity in the stomach is unstable and difficult because of respiratory fluctuations. If the target lesion is small, it is extremely difficult to identify the lesion with pCLE, so it is impossible to completely eliminate the

influence of WLE interpretation from pCLE diagnosis for gastric cancer. For this reason, reviewers were not blinded to WLE findings. Owing to these disadvantages, we believe that CLE will not replace forceps biopsy in endoscopy examinations. However, pCLE could assist biopsy-based histopathological diagnosis. If a lesion is highly likely to be neoplastic according to WLE findings and considered for endoscopic resection, pCLE could replace preoperative biopsy; this would avoid creation of biopsy-induced fibrosis, which can adversely affect endoscopic resection.

In conclusion, the results of our study show that reviewers' clinical backgrounds influence the interpretation of pCLE used to diagnose superficial gastric lesions. Japanese reviewers achieved good results during pCLE interpretation, possibly because of the high prevalence of gastric cancer in Japan. We surmise that expertise in WLE is essential to achieving accurate diagnosis by pCLE.

REFERENCES

1 Kiesslich R, Burg J, Vieth M, et al. Confocal laser endoscopy for diagnosing intraepithelial neoplasias and colorectal cancer in vivo. Gastroenterology 2004; 127: 706-713

2 Kiesslich R, Gossner L, Goetz M, et al. In vivo histology of Barrett's esophagus and associated neoplasia by confocal laser endomicroscopy. Clin Gastroenterol Hepatol 2006; 4: 979-987

3 De Palma GD. Confocal laser endomicroscopy in the "in vivo" histological diagnosis of the gastrointestinal tract. World J Gastroenterol 2009; 15: 5770-5775

4 Neumann H, Kiesslich R, Wallace MB, et al. Confocal laser endomicroscopy: technical advances and clinical applications. Gastroenterology 2010; 139: 388-92, 392.e1-2

5 Ohya TR, Sumiyama K, Takahashi-Fujigasaki J, et al. In vivo histologic imaging of the muscularis propria and myenteric neurons with probe-based confocal laser endomicroscopy in porcine models (with videos). Gastrointest Endosc 2012; 75:

405-410

6 Sumiyama K, Kiesslich R, Ohya TR, et al. In vivo imaging of enteric neuronal networks in humans using confocal laser endomicroscopy. Gastroenterology 2012; 143: 1152-1153

7 Wang KK, Carr-Locke DL, Singh SK, et al. Use of probe-based confocal laser endomicroscopy (pCLE) in gastrointestinal applications. A consensus report based on clinical evidence. United European Gastroenterol J 2015; 3: 230-254

8 Canto MI, Anandasabapathy S, Brugge W, et al. In vivo endomicroscopy improves detection of Barrett's esophagus-related neoplasia: a multicenter international randomized controlled trial (with video). Gastrointest Endosc 2014; 79: 211-221

9 Bessissow T, Bisschops R. Advanced endoscopic imaging for dysplasia surveillance in ulcerative colitis. Expert Rev Gastroenterol Hepatol 2013; 7: 57-67

10 Kitabatake S, Niwa Y, Miyahara R, et al. Confocal endomicroscopy for the diagnosis of gastric cancer in vivo. Endoscopy 2006; 38: 1110-1114

11 Zhang JN, Li YQ, Zhao YA, et al. Classification of gastric pit patterns by

confocal endomicroscopy. Gastrointest Endosc 2008; 67: 843-853

12 Li Z, Zuo XL, Yu T, et al. Confocal laser endomicroscopy for in vivo detection of gastric intestinal metaplasia: a randomized controlled trial. Endoscopy 2014; 46: 282-290

13 Peter S, Council L, Bang JY, et al. Poor agreement between endoscopists and gastrointestinal pathologists for the interpretation of probe-based confocal laser endomicroscopy findings. World J Gastroenterol 2014; 20: 17993-18000

14 Bajbouj M, Vieth M, Rosch T, et al. Probe-based confocal laser endomicroscopy compared with standard four-quadrant biopsy for evaluation of neoplasia in Barrett's esophagus. Endoscopy 2010; 42: 435-440

15 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: 188-193

16 Pohl H, Rosch T, Vieth M, et al. Miniprobe confocal laser microscopy for the detection of invisible neoplasia in patients with Barrett's oesophagus. Gut 2008; 57: 1648-1653

17 Neumann H, Neurath MF, Mudter J. New endoscopic approaches in IBD. World J Gastroenterol 2011; 17: 63-68

18 Neumann H, Vieth M, Atreya R, et al. Prospective evaluation of the learning curve of confocal laser endomicroscopy in patients with IBD. Histol Histopathol 2011; 26: 867-872

19 Neumann H, Vieth M, Atreya R, et al. Assessment of Crohn's disease activity by confocal laser endomicroscopy. Inflamm Bowel Dis 2012; 18: 2261-2269

20 Neumann H, Langner C, Neurath MF, et al. Confocal Laser Endomicroscopy for Diagnosis of Barrett's Esophagus. Front Oncol 2012; 2: 42

21 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

22 Kakeji Y, Yamaguchi S, Yoshida D, et al. Development and assessment of morphologic criteria for diagnosing gastric cancer using confocal endomicroscopy: an ex vivo and in vivo study. Endoscopy 2006; 38: 886-890

23 Li Z, Zuo XL, Li CQ, et al. In vivo molecular imaging of gastric cancer by targeting MG7 antigen with confocal laser endomicroscopy. Endoscopy 2013; 45: 79-85

24 Menon S, Trudgill N. How commonly is upper gastrointestinal cancer missed at endoscopy? A meta-analysis. Endosc Int Open 2014; 2: E46-50

25 Suzuki H, Oda I, Abe S, et al. High rate of 5-year survival among patients with early gastric cancer undergoing curative endoscopic submucosal dissection. Gastric Cancer 2015

26 Maruyama K, Katai H, -. Surgical treatment of gastric cancer in Japan, trend from standardization to individualization. Chirurgia (Bucur) 2014; 109: 722-730

27 Wallace M, Lauwers GY, Chen Y, et al. Miami classification for probe-based confocal laser endomicroscopy. Endoscopy 2011; 43: 882-891

28 Schlemper RJ, Riddell RH, Kato Y, et al. The Vienna classification of gastrointestinal epithelial neoplasia. Gut 2000; 47: 251-255

29 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: 556-560

30 Hirayama Y, Kawai T, Otaki J, et al. Prevalence of Helicobacter pylori infection with healthy subjects in Japan. J Gastroenterol Hepatol 2014; 29 Suppl 4: 16-19

31 Bok GH, Jeon SR, Cho JY, et al. The accuracy of probe-based confocal endomicroscopy versus conventional endoscopic biopsies for the diagnosis of superficial gastric neoplasia (with videos). Gastrointest Endosc 2013; 77: 899-908

FIGURE LEGENDS

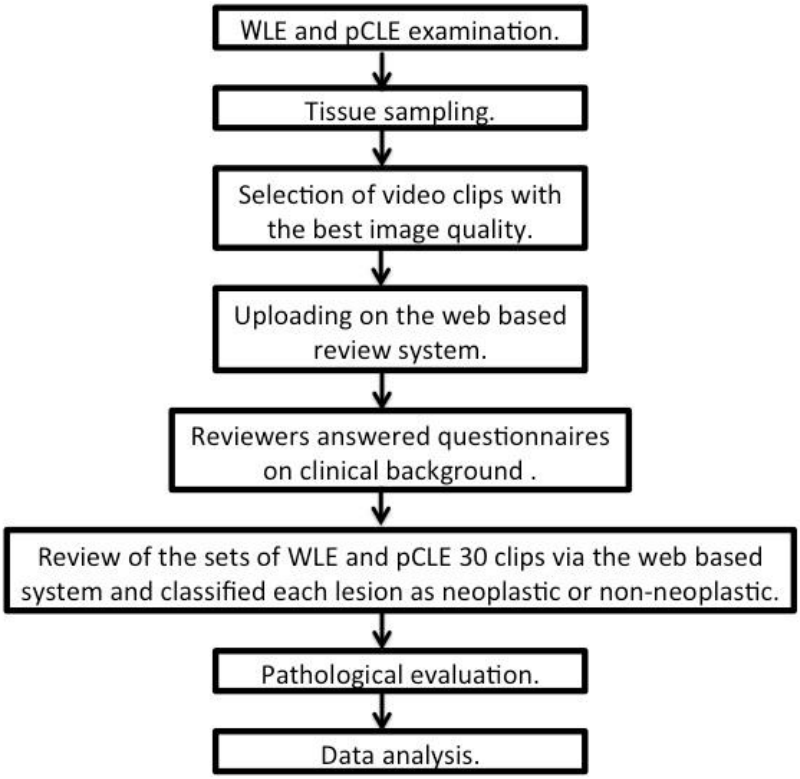
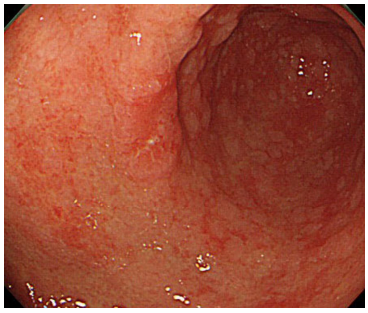
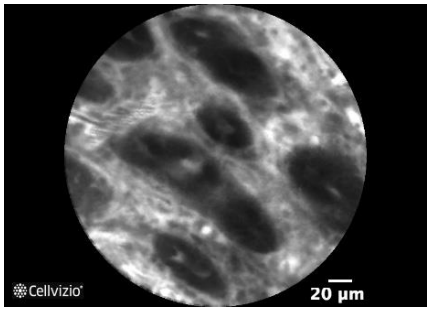


Figure 1. Flow chart of the study protocol.

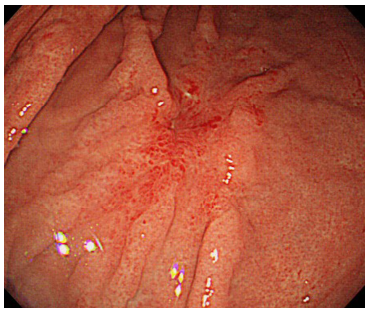
Figure 2. Representative WLE and pCLE images of superficial lesions.



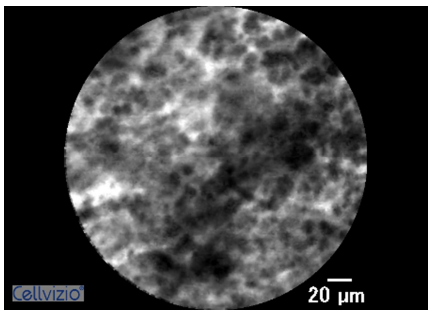
(a) WLE image of a moderately differentiated adenocarcinoma.



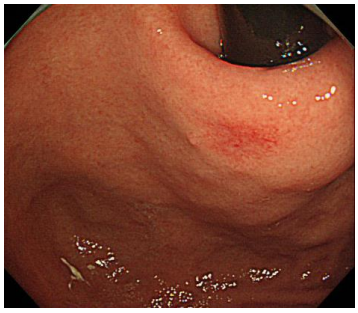
(b) pCLE image of the lesion shown in (a).



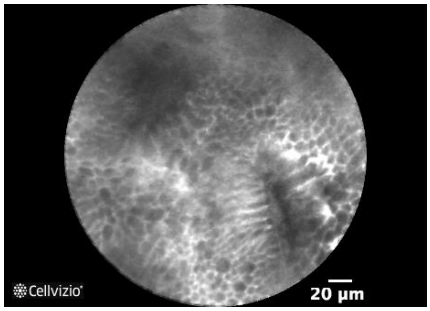
(c) WLE image of a poorly differentiated adenocarcinoma.



(d) pCLE image of the lesion shown in (c).



(e) WLE image of a non-neoplastic lesion.



(f) pCLE image of the lesion shown in (e).

VIDEO LEGEND

Video. Representative WLE and pCLE clips of a moderately differentiated adenocarcinoma.

TABLES

Table 1: Questions about reviewers' backgrounds and possible responses

Question	**Possible responses**
Country	Germany / Japan
Time since graduation from medical school (years)	<5 / 5–10 / >10
Time since finishing gastroenterology fellowship (years)	<5 / 5–10 / >10
Duration of GI endoscopy practice (years)	<5 / 5–10 / >10
Specialized field	GI physician, GI surgeon / Pathologist
No. of cases of GI flexible endoscopy	<1000 / ≥1000
No. of cases of confocal endoscopy	0 / 1–19 / 20–50 / 51–100 / >100
Pathology fellowship or training in GI pathology	Yes / No
Duration of pathology training (years)	<1 / 1–3 / >3

Abbreviation: GI, gastrointestinal

Table 2; Lesions inclued in the study

	Patient characteristics		Lesion characteristics				
No.	**Age (years)**	**Gender**	**Location**		**Morphological type**	**Size (mm)**	**Histopathological diagnosis**
1	60	Female	L	l.c.	Ⅱc	5	Ca
2	60	Female	L	p.w.	Ⅱc	5	Ca
3	62	Male	L	a.w.	Ⅱc	6	Ca
4	64	Male	L	l.c.	Ⅱa	8	HGD
5	52	Male	L	a.w.	Ⅱa	20	LGD
6	56	Female	L	l.c.	Ⅱa	5	LGD
7	71	Female	L	a.w.	Ⅱa＋Ⅱc	35	HGD
8	60	Male	L	l.c.	Ⅱc	10	Ca
9	62	Male	L	g.c.	Ⅱc	8	Ca
10	62	Female	L	a.w.		3	No neoplasia
11	55	Male	L	g.c.	Ⅱc	15	Ca
12	47	Male	L	a.w.		5	No neoplasia
13	63	Female	L	g.c.	Ⅱc	25	Ca
14	63	Female	L	p.w.		5	No neoplasia
15	71	Male	L	a.w.		15	No neoplasia
16	71	Male	L	a.w.	Ⅱc	25	Ca
17	71	Male	L	l.c.	Ⅱa+Ⅱc	5	Ca
18	71	Male	L	p.w.		5	No neoplasia
19	71	Male	L	a.w.		3	No neoplasia
20	59	Male	L	p.w.		10	No neoplasia
21	59	Male	L	g.c.		5	No neoplasia
22	59	Male	L	g.c.		10	No neoplasia
23	64	Male	L	g.c.	Ⅱc	5	Ca
24	73	Male	M	l.c.	Ⅱb	30	Ca
25	73	Male	M	l.c.	Ⅱc+Ⅱb	10	Ca
26	68	Female	U	p.w.	Ⅱc	35	Ca
27	68	Female	U	l.c.		5	No neoplasia
28	68	Female	U	g.c.	Ⅱc	15	LGD
29	71	Male	U	l.c.		5	No neoplasia
30	49	Male	U	p.w.		5	No neoplasia

Abbreviations: L, lower-third of the stomach; M, middle-third of the stomach; U, upper-third of the stomach; l.c., lesser curvature; g.c., greater curvature; a.w., anterior wall; p.w., posterior wall; Ca, carcinoma; HGD, high-grade dysplasia; LGD, low-grade dysplasia

Table 3: Diagnostic yield of WLE in each group of reviewers

Clinical background	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	LR+ (95% CI)	LR- (95% CI)
Overall (n = 39)	58.97% (55.23%–62.64%)	75.64% (71.49%–79.46%)	78.40% (74.65%–81.85%)	55.14% (51.20%–59.03%)	2.42 (2.04–2.87)	0.54 (0.49–0.60)
With CLE experience (n = 7)	53.17% (44.08%–62.12%)	80.95% (70.92%–88.70%)	80.72% (70.59%–88.56%)	53.54% (44.48%–62.44%)	2.79 (1.74–4.47)	0.58 (0.47–0.72)
Without CLE experience (n = 32)	60.24% (56.12%–64.27%)	74.48% (69.81%–78.77%)	77.98% (73.84%–81.74%)	55.53% (51.12%–59.88%)	2.36 (1.97–2.84)	0.53 (0.48–0.60)
GI physicians and surgeons (n = 33)	58.59% (54.51%–62.58%)	79.29% (74.96%–83.18%)	80.93% (76.89%–84.54%)	56.07% (51.85%–60.23%)	2.83 (2.31–3.47)	0.52 (0.47–0.58)
Pathologists (n = 6)	61.11% (51.25%–70.34%)	55.56% (43.36%–67.28%)	67.35% (57.13%–76.48%)	48.78% (37.58%–60.08%)	1.38 (1.02–1.85)	0.70 (0.51–0.96)
German (n = 7)	52.38% (43.30%–61.35%)	60.71% (49.45%–71.20%)	66.67% (56.48%–75.82%)	45.95% (36.45%–55.67%)	1.33 (0.97–1.82)	0.78 (0.61–1.01)
Japanese (n = 32)	60.42% (56.29%–64.43%)	78.91% (74.48%–82.88%)	81.12% (77.09%–84.71%)	57.06% (52.73%–61.32%)	2.86 (2.33–3.51)	0.50 (0.45–0.56)
Expert endoscopists with pathology training (n = 14)	58.33% (51.98%–64.49%)	80.95% (74.19%–86.59%)	82.12% (75.71%–87.44%)	56.43% (49.92%–62.79%)	3.06 (2.20–4.25)	0.52 (0.44–0.61)
Expert endoscopists without pathology training (n = 17)	60.13% (54.50%–65.66%)	78.92% (72.68%–84.31%)	81.06% (75.34%–85.94%)	56.89% (50.90%–62.74%)	2.85 (2.15–3.78)	0.51 (0.43–0.59)

Abbreviations: WLE, white-light endoscopy; CLE, confocal laser endomicroscopy; CI, confidence interval; GI, gastrointestinal; PPV, positive predictive value; NPV, negative predictive value; LR+, positive likelihood ratio; LR-, negative likelihood ratio

Table 4: Diagnostic yield of WLE + pCLE in each group of reviewers

Clinical background	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	LR+ (95% CI)	LR- (95% CI)
Overall (n = 39)	70.37% (66.84%–73.73%)	79.27% (75.31%–82.86%)	83.59% (80.35%–86.48%)	64.08% (60.02%–67.99%)	3.40 (2.83–4.08)	0.37 (0.33–0.42)
With CLE experience (n = 7)	69.05% (60.20%–76.98%)	85.71% (76.38%–92.39%)	87.88% (79.78%–93.58%)	64.86% (55.23%–73.69%)	4.83 (2.83–8.27)	0.36 (0.27–0.48)
Without CLE experience (n = 32)	70.66% (66.76%–74.35%)	77.86% (73.38%–81.92%)	82.72% (79.09%–85.96%)	63.89% (59.35%–68.25%)	3.19 (2.63–3.88)	0.38 (0.33–0.43)
GI physicians and surgeons (n = 33)	71.55% (67.74%–75.15%)	81.82% (77.66%–85.49%)	85.51% (82.11%–88.49%)	65.72% (61.34%–69.91%)	3.94 (3.17–4.88)	0.35 (0.30–0.40)
Pathologists (n = 6)	63.89% (54.08%–72.91%)	65.28% (53.14%–76.12%)	73.40% (63.29%–81.99%)	54.65% (43.55%–65.42%)	1.84 (1.30–2.60)	0.55 (0.41–0.75)
German (n = 7)	59.52% (50.42%–68.17%)	71.43% (60.53%–80.76%)	75.67% (66.11%–83.81%)	54.05% (44.33%–63.55%)	2.08 (1.44–3.01)	0.57 (0.44–0.73)
Japanese (n = 32)	72.74% (68.91%–76.34%)	80.99% (76.70%–84.79%)	85.16% (81.71%–88.19%)	66.45% (61.97%–70.72%)	3.83 (3.09–4.73)	0.34 (0.29–0.39)
Expert endoscopists with pathology training (n = 14)	69.44% (63.35%–75.07%)	79.17% (72.24%–85.04%)	83.33% (77.59%–88.11%)	63.33% (56.43%–69.86%)	3.33 (2.45–4.53)	0.39 (0.32–0.47)
Expert endoscopists without pathology training (n = 17)	73.86% (68.55%–78.69%)	82.35% (76.42%–87.32%)	86.26% (81.49%–90.19%)	67.74% (61.54%–73.52%)	4.19 (3.09–5.67)	0.32 (0.26–0.39)

Abbreviations: WLE, white-light endoscopy; CLE, confocal laser endomicroscopy; CI, confidence interval; GI, gastrointestinal; PPV, positive predictive value; NPV, negative predictive value; LR+, positive likelihood ratio; LR-, negative likelihood ratio

Table 5: Accuracy of the differential diagnosis in each group of reviewers

	Accuracy (95% CI)		
Clinical background	**WLE**	**WLE + pCLE**	***P***
Overall (n = 39)	65.64% (62.84%–68.36%)	73.93% (71.3%–76.42%)	.0002
With CLE experience (n = 7)	64.29% (57.4%–70.76%)	75.71% (69.34%–81.35%)	.0195
Without CLE experience (n = 32)	65.94% (62.84%–68.93%)	73.54% (70.63%–76.31%)	.0028
GI physicians and surgeons (n = 33)	66.87% (63.84%–69.80%)	75.66% (72.86%–78.30%)	< .01
Pathologists (n = 6)	58.89% (51.32%–66.15%)	64.44% (56.90%–71.42%)	.6358
German (n = 7)	55.71% (48.72%–62.55%)	64.9% (57.40%–70.76%)	1.000
Japanese (n = 32)	67.81% (64.75%–70.76%)	73.21% (73.21%–78.71%)	< .01
Expert endoscopists with pathology training (n = 14)	67.38% (62.66%–71.84%)	73.33% (68.83%–77.50%)	.002
Expert endoscopists without pathology training (n = 17)	67.64% (63.39%–71.69%)	77.25% (73.37%–80.82%)	.002

Abbreviations: WLE, white-light endoscopy; pCLE, probe-based confocal laser endomicroscopy; CLE, confocal laser endomicroscopy; CI, confidence interval; GI, gastrointestinal

Table 6: Levels of inter-observer agreement between reviewers in each clinical background group

	Kappa value	
Clinical background	**WLE**	**WLE + pCLE**
Overall (n = 39)	0.30	0.34
With CLE experience (n = 7)	0.25	0.46
Without CLE experience (n = 32)	0.31	0.32
GI physicians and surgeons (n = 33)	0.33	0.39
Pathologists (n = 6)	0.16	0.09
German (n = 7)	0.18	0.15
Japanese (n = 32)	0.35	0.39
Expert endoscopists with pathology training (n = 14)	0.30	0.35
Expert endoscopists without pathology training (n = 17)	0.36	0.42

Abbreviations: WLE, white-light endoscopy; pCLE, probe-based confocal laser endomicroscopy; CLE, confocal laser endomicroscopy; GI, gastrointestinal.

Table 7. Results of multilevel logistic regression analysis used to identify variables affecting the correct diagnosis based on WLE + pCLE

Variables	**OR (95% CI)**	***P***
German	0.65 (0.42–0.99)	.05
Pathologist	0.78 (0.43–1.42)	.42
Endoscopy experience	1.00 (0.62–1.64)	.99
CLE experience	0.86 (0.76–0.98)	.02
Pathology training	0.86 (0.62–1.19)	.36

Abbreviations: WLE, white-light endoscopy; pCLE, probe-based confocal laser endomicroscopy; CLE, confocal laser endomicroscopy; OR, odds ratio; CI, confidence interval

Table 8. Results of multilevel logistic regression analysis used to identify variables affecting the correct diagnosis based on WLE and WLE + pCLE.

Variable	**OR (95% CI)**	***P***
German	0.61 (0.41–0.91)	.02
WLE + pCLE	1.51 (1.26–1.81)	< .01
Pathologist	0.88 (0.50–1.56)	.67
Endoscopy experience	1.10 (0.70–1.72)	.69
CLE experience	0.84 (0.74–0.94)	< .01
Pathology training	0.96 (0.71–1.30)	.79

Abbreviations: WLE, white-light endoscopy; pCLE, probe-based confocal laser endomicroscopy; CLE, confocal laser endomicroscopy; OR, odds ratio; CI, confidence interval