



Artificial intelligence for endoscopy in inflammatory bowel disease

Kento Takenaka¹, Ami Kawamoto¹, Ryuichi Okamoto¹, Mamoru Watanabe², Kazuo Ohtsuka³

¹Department of Gastroenterology and Hepatology, ²TMDU Advanced Research Institute, and ³Endoscopic Unit, Tokyo Medical and Dental University, Tokyo, Japan

Inflammatory bowel disease (IBD), with its 2 subtypes, Crohn's disease and ulcerative colitis, is a complex chronic condition. A precise definition of disease activity and appropriate drug management greatly improve the clinical course while minimizing the risk or cost. Artificial intelligence (AI) has been used in several medical diseases or situations. Herein, we provide an overview of AI for endoscopy in IBD. We discuss how AI can improve clinical practice and how some components have already begun to shape our knowledge. There may be a time when we can use AI in clinical practice. As AI systems contribute to the exact diagnosis and treatment of human disease, we should continue to learn best practices in health care in the field of IBD. (**Intest Res 2022;20:165-170**)

Key Words: Colitis, ulcerative; Crohn disease; Mucosal healing

INTRODUCTION

Inflammatory bowel disease (IBD), with its 2 subtypes, Crohn's disease (CD) and ulcerative colitis (UC), is a complex chronic condition with a wide range of contributing factors. A precise definition of disease activity and appropriate drug management greatly improve the clinical course while minimizing the risk or cost. Artificial intelligence (AI) has been used in several medical diseases or situations. Herein, we provide an overview of AI for endoscopy in IBD. We discuss how AI can improve clinical practice in IBD and how some components have already begun to shape our knowledge.

NEED OF AI FOR ENDOSCOPY

The evaluation of endoscopic inflammation, characterization

of lesions, and mucosal healing assessment are essential for proper IBD management. Endoscopic remission is associated with improved long-term outcomes and is recommended as a treatment target.¹ Endoscopic scoring has important implications for clinical trial outcomes and routine practice care.² However, endoscopic assessment of inflammation is highly subjective, and interobserver and intraobserver variability in evaluating inflamed mucosa is high.³ Recent evidence has suggested that histologic remission is associated with an independent benefit for long-term outcomes,⁴ and the need for histological evaluation of the colonic mucosa has also been emphasized (especially for UC patients).⁵ Healthcare providers should perform lower endoscopy with biopsies and visually interpret the histological parameters of inflammation to assess these outcomes. Image recognition, particularly deep learning, is a major AI application that holds great promise in assisting medical imaging. Computer-aided diagnosis (CAD) is becoming an increasingly popular means of addressing human error. CAD for IBD endoscopy allows assessments with less bias and more objective interpretation.

Received June 8, 2021. Accepted August 25, 2021.

Correspondence to Kento Takenaka, Department of Gastroenterology and Hepatology, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8519, Japan. Tel: +81-3-5803-5877, Fax: +81-3-5803-0268, E-mail: ktakenaka.gast@tmd.ac.jp

CAD SYSTEM FOR UC

The use of the CAD system for UC has been reported from several institutions (Table 1).

A retrospective analysis by Ozawa et al.⁶ reported the construction of a CAD system evaluated by tagging a dataset of standard endoscopic UC images from patients. The trained CAD identified normal mucosa, a Mayo endoscopic subscore (MES) of 0, and mucosal healing, MES 0–1. It showed excellent performance with the area under a receiver operating characteristic curve (AUROC) values of 0.86 and 0.98 for differentiating MES 0 from 1 to 3 and MES 0–1 from 2 to 3, respectively.⁷

Maeda et al.⁸ developed a CAD system that uses endocytoscopy to predict persistent histologic inflammation in UC patients. Endocytoscopy is performed with a 520-fold ultra-magnifying contact light microscope comparable with other advanced technologies to predict histological severity. This CAD system showed a good prediction of histological activity (defined as a Geboes score⁹ <3.0) with a sensitivity, specificity, and accuracy of 74%, 97%, and 91%, respectively. The authors concluded that this system could contribute to fully automated identification of persistent histological inflammation associated with UC. However, it is important to point out that endocytoscopy is not generally used in clinical practice.

Stidham et al.¹⁰ investigated grading the endoscopic severity of UC and applied it to full-motion video from standard colonoscopy. Based on data from Michigan's endoscopic imaging

database (16,514 colonoscopic images from 3,082 UC patients), a CAD system was constructed to categorize images into 2 groups: endoscopic remission (defined as an MES 0–1) and moderate-to-severe disease (MES 2–3). The results showed that it could distinguish endoscopic remission from active disease with an AUROC of 0.97, a sensitivity of 83%, a specificity of 96%, a positive predictive value of 87%, and a negative predictive value of 94%. This CAD system was then applied to entire full-motion colonoscopy videos in a recent pilot study.¹¹ Non-informative images were identified by several qualitative characteristics (proximity to tissue, light reflection, debris and blur, and motion blur), and whole-video Mayo endoscopic scores were estimated. The validity was evaluated on a developmental set of high-resolution videos (51 videos) and a multicenter clinical trial set (264 videos). Fully automated methods correctly predicted MES in 78% for high-resolution videos and 83% for external clinical trial videos, respectively. A striking aspect of this study was that the researchers conducted complete external validation using standard colonoscopy videos to mimic real-world application.

Bossuyt et al.¹² developed CAD to output a red density (RD) score based on images from a prototype endoscope. Their algorithm is based on the integration of pixel color data along with vessel pattern detection. The results showed that the RD score correlated with Robarts histological index¹³ ($r=0.74$, $P<0.01$), the MES ($r=0.76$, $P<0.01$), and the Ulcerative Colitis Endoscopic Index of Severity (UCEIS)¹⁴ ($r=0.74$, $P<0.01$). Because

Table 1. A Summary of Computer-Aided Characterization for Endoscopy on Ulcerative Colitis

Author (year)	Study design	Training samples	Test samples	Type of colonoscopy	Outcome	Histology prediction
Maeda et al. (2019) ⁸	Retrospective	Images (87 pt)	Images (100 pt)	Endocytoscopy (ultra-magnifying endoscope)	Predicting histologic remission	Yes
Ozawa et al. (2019) ⁶	Retrospective	Images (841 pt)	Images (114 pt)	Standard colonoscopy	Identifying MES 0 or MES 0-1	No
Stidham et al. (2019) ¹⁰	Retrospective	Images (2,778 pt)	Images (304 pt) Videos (30 pt)	Standard colonoscopy	Identifying MES 0-1	No
Bossuyt et al. (2020) ¹²	Prospective	Images (29 pt)	Images (10 pt)	Prototype endoscope	Determining red density score which correlated with endoscopic and histologic scores	Yes
Takenaka et al. (2020) ¹⁵	Prospective	Images (2,012 pt)	Images (875 pt)	Standard colonoscopy	Determining UCEIS	Yes
Gottlieb et al. (2021) ¹⁷	Retrospective (post-hoc)	Videos (80% of 249 pt)	Videos (20% of 249 pt)	Standard colonoscopy	Determining MES and UCEIS	No
Yao et al. (2021) ¹¹	Prospective	-	Videos (51 pt)	Standard colonoscopy	Determining MES	No

pt, patients; MES, Mayo endoscopic score; UCEIS, Ulcerative Colitis Endoscopic Index of Severity.

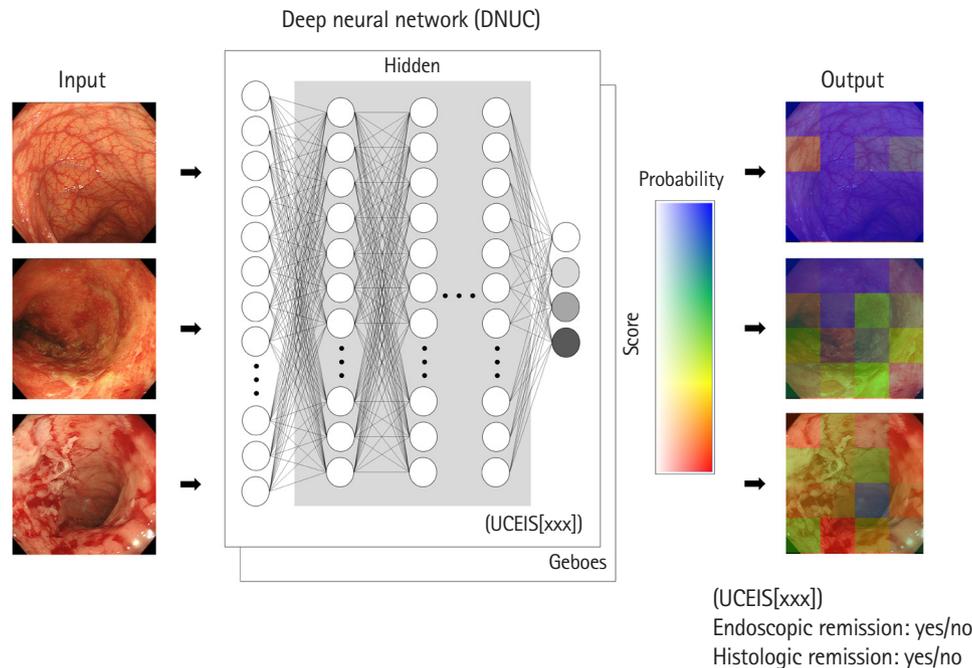


Fig. 1. Example of a computer-aided diagnosis system. Takenaka et al.¹⁵ constructed a deep neural network to evaluate ulcerative colitis (DNUC) from endoscopic images. This system determined the Ulcerative Colitis Endoscopic Index of Severity (UCEIS) score and histological activity. This figure is published with the permission of the author.

these results showed that CAD could determine not only endoscopic findings but also histology, they concluded that the RD system is a novel modality that provides an objective computer-based score that accurately assesses UC disease activity.

Takenaka et al.¹⁵ constructed a deep neural network for the evaluation of UC (called DNUC) using 40,758 colonoscopy images tagged with 6,885 biopsy results (Fig. 1). The accuracy of this CAD system was validated in a prospective study of 875 patients with UC who underwent standard colonoscopy with 4,187 endoscopic images and 4,104 biopsy specimens. This system determined the UCEIS score and the Geboes score. The results demonstrated that the accuracy of identifying endoscopic remission (defined as a UCEIS of 0) was 90%, and identifying histological remission (defined as a Geboes score of <3.1) was 93%. Also, the correlation between AI and IBD experts for scoring the UCEIS was high at 0.917. Furthermore, the prognostic value of the DNUC was evaluated in a prospective cohort study.¹⁶ Mucosal healing (a combination of endoscopic and histological remission) identified by CAD was associated with a significantly lower risk of worse prognosis ($P < 0.001$ for hospitalization, colectomy, steroid use, and clinical relapse). The prognostic value was calculated with a hazard ratio, and the differences between the results obtained by the experts and CAD were not statistically significant (hospitaliza-

tion, $P = 0.367$; colectomy, $P = 0.693$; steroid use, $P = 0.851$; and relapse, $P = 0.758$). This longitudinal study supports the consistent evaluation of mucosal healing with CAD and its future use. A multicenter prospective study is needed to confirm this system's accuracy because the DNUC algorithm was constructed with data from a single center and still images.

Gottlieb et al.¹⁷ investigated whether a CAD system for video colonoscopy could replace central reading in a recent *post hoc* study. They developed CAD to determine MES and UCEIS scores using full-length endoscopic video from a phase 2 trial of mirikizumab. The developed model's agreement was excellent, with a quadratic weighted kappa of 0.844 for MES and 0.855 for UCEIS. These results support that the CAD system can be trained to predict UC severity levels from full-length endoscopy videos.

CAD SYSTEM FOR CD

The treat-to-target approach has also emerged as an important treatment strategy in patients with CD. One of the gold-standard targets for inflammation improvement has been endoscopic remission based on ileocolonoscopy evaluation,¹ or balloon-assisted enteroscopy for ileal type.¹⁸ However, the morphologic and anatomic variation typical of CD poses problems for current image classification technologies using AI. As a re-

sult, work replicating common endoscopic scores such as Simple Endoscopic Score for Crohn's Disease and Crohn's Disease Index of Severity has been limited. In contrast, current AI-based image classification is proving useful for aiding the detection of small bowel ulcerations using video capsule endoscopy (VCE). VCE is an accurate clinical tool for diagnosis and monitoring of CD, and small bowel evaluation with capsule endoscopy (CE) is recommended in all newly diagnosed CD patients and in patients with established CD with clinical exacerbation or unexplained symptoms.¹⁹ The diagnostic yield of VCE is similar to cross-sectional imaging for detection of active endoscopic inflammation in established CD.²⁰ Active endoscopic inflammation in the small bowel is frequently detected even in patients in clinical remission and significantly impacts on relapse-free survival.²¹ Klang et al.²² developed deep learning technology to provide accurate and fast automated detection of mucosal ulcers on VCE. They also reported that deep neural networks were highly accurate in the detection of CD-related strictures on CE, and accurately separated strictures from ulcers across the severity range.²³ Barash et al.²⁴ reported AI achieved a high accuracy in detecting severe CD ulcerations and concluded that AI-assisted CE readings in patients with CD can potentially facilitate and improve diagnosis and monitoring in these patients. Ding et al.²⁵ showed that automated lesion detection methods reduced mean VCE review times from 96.6 minutes by 5.9 minutes when using computer assisted reading with no differences in sensitivity for disease findings. Although the heterogeneity of CD presents challenges that will require further technologic developments, current methods may still prove useful in easing the time burden and improving sensitivity for reviewing VCE.

FUTURE ASPECTS OF CAD

AI has begun to demonstrate expert-level judgment using cleaned

and curated data. It is now beginning to show promise for understanding endoscopic evaluation. The progress of CAD is remarkable, and we think that there are 2 important positions for using CAD for the endoscopic evaluation for IBD in the future (Fig. 2). First, AI always outputs the same result from the same images or videos, enabling objective and consistent endoscopic evaluation. This standardization would be very useful not only for clinical practice but also for central reading in clinical trials or gastroenterologist training. A precise and detailed real-time assessment of the mucosa has become more important than ever for the medical management of IBD patients.²⁶ Although endoscopic outcomes are important endpoints in clinical trials or research, these assessments are subjective, and central blinded reading is necessary. In addition, local site investigators tend to systematically overscore baseline endoscopic severity compared with remote investigators.²⁷ Endoscopic score reproducibility, reliability, and objectivity have improved with central reading by experienced reviewers.² However, central reading is time-consuming and cost-intensive, with uncertain applications for routine care. Immediate objective blinded CAD assessments would solve this limitation and help advance the routine high-quality interpretation of endoscopic scoring for incorporation as treatment targets. The notion that CAD may be used to train future gastroenterology fellows is interesting. Current trainees are expected to achieve competency in precise interpretation of endoscopic activity when performing endoscopies for IBD. However, their experiences are largely subjective and require that the supervisor have a robust understanding and an accurate ability to score endoscopy. A standardized, accurate, and objective assessment of mucosal disease activity using AI can verify their endoscopic score interpretation in real-time and identify and strengthen knowledge gaps. The second position is that CAD can benefit disease management by improving the cost-effectiveness of daily practice patterns. Since CAD provides a con-

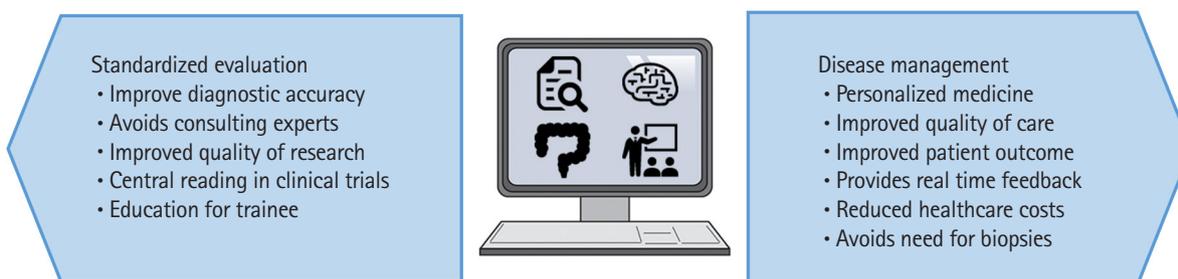


Fig. 2. Aspect of artificial intelligence for endoscopy in inflammatory bowel disease. In the future, great advantages are expected for standardized evaluation or disease management.

sistent endoscopic evaluation similar to IBD experts, we can avoid the need to consult experts in community practice settings. In addition, histological assessments in UC are now important targets, but they require additional time, processing, and interpretation, limiting real-time decision making. Several CAD systems have achieved prediction of histological inflammation only from endoscopic images and showed the potential of reducing the need for biopsies. This advantage would provide a cost-benefit by obviating the need to collect and process specimens and avoid specialized pathologists' requirements.

AI can revolutionize how we practice medicine; however, there are several barriers to overcome before general use in routine clinical care. First, the output process of AI is extremely complex and is beyond the scope of human understanding. Thus, physicians must interpret the results with caution. However, we believe that the assessment of longitudinal responsiveness could facilitate its use in clinical settings.¹⁶ Second, a highly diverse and comprehensive dataset must be input to train AI, which incorporates several disease phenotypes, treatment exposures, and image quality. Further prospective validation in alternative clinical practice datasets and endoscopic devices is needed to ensure the generalizability of developed CAD. Third, we reviewed the CAD for UC, but the corresponding evidence on CD is limited at present. CD is a disease with transmural inflammation, and extraintestinal disease is also an important complication. The endoscope's position in treat-to-target strategy in CD is controllable, and advances in AI, including cross-sectional images, are expected in the future. Finally, CAD could predict histological activity from endoscopic images or videos, but a detailed evaluation of histology or grading histological scores was impossible. Therefore, we consider that the gold standard of histology needs to be performed with mucosal biopsies, and their importance should not be ignored. A CAD approach does not replace the need for dysplasia surveillance and routine surveillance biopsies in UC patients. Future iterations incorporating previous AI systems to detect adenomas and dysplasia combined with histology would prove to be of great value in overcoming this final hurdle.

CONCLUSIONS

There may be a time when we can use AI in clinical practice. As AI systems contribute to the exact diagnosis and treatment of human disease, we should continue to learn best practices in health care in the field of IBD.

ADDITIONAL INFORMATION

Funding Source

The authors received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest

Watanabe M is an editorial board member of the journal but was not involved in the peer reviewer selection, evaluation, or decision process of this article. No other potential conflicts of interest relevant to this article were reported.

Author Contribution

Conceptualization: Takenaka K. Data curation: Takenaka K. Investigation: Takenaka K. Methodology: Takenaka K. Project administration: Takenaka K. Supervision: Okamoto R, Watanabe M, Ohtsuka K. Visualization: Takenaka K. Writing - original draft: Takenaka K. Writing - review & editing: Kawamoto A. Approval of final manuscript: all authors.

ORCID

Takenaka K	https://orcid.org/0000-0002-7861-8383
Kawamoto A	https://orcid.org/0000-0002-8549-4999
Okamoto R	https://orcid.org/0000-0002-7047-571X
Watanabe M	https://orcid.org/0000-0002-5475-9544
Ohtsuka K	https://orcid.org/0000-0002-2409-9709

REFERENCES

1. Turner D, Ricciuto A, Lewis A, et al. STRIDE-II: an update on the Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE) Initiative of the International Organization for the Study of IBD (IOIBD): determining therapeutic goals for treat-to-target strategies in IBD. *Gastroenterology* 2021;160:1570-1583.
2. Feagan BG, Sandborn WJ, D'Haens G, et al. The role of centralized reading of endoscopy in a randomized controlled trial of mesalazine for ulcerative colitis. *Gastroenterology* 2013;145:149-157.
3. Osada T, Ohkusa T, Yokoyama T, et al. Comparison of several activity indices for the evaluation of endoscopic activity in UC: inter- and intraobserver consistency. *Inflamm Bowel Dis* 2010;16:192-197.
4. Kaneshiro M, Takenaka K, Suzuki K, et al. Pancolonic endoscopic and histologic evaluation for relapse prediction in patients with ulcerative colitis in clinical remission. *Aliment Phar*

- macol Ther 2021;53:900-907.
5. Cushing KC, Tan W, Alpers DH, Deshpande V, Ananthakrishnan AN. Complete histologic normalisation is associated with reduced risk of relapse among patients with ulcerative colitis in complete endoscopic remission. *Aliment Pharmacol Ther* 2020;51:347-355.
 6. Ozawa T, Ishihara S, Fujishiro M, et al. Novel computer-assisted diagnosis system for endoscopic disease activity in patients with ulcerative colitis. *Gastrointest Endosc* 2019;89:416-421.
 7. Higgins PD, Schwartz M, Mapili J, Krokos I, Leung J, Zimmermann EM. Patient defined dichotomous end points for remission and clinical improvement in ulcerative colitis. *Gut* 2005;54:782-788.
 8. Maeda Y, Kudo SE, Mori Y, et al. Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video). *Gastrointest Endosc* 2019;89:408-415.
 9. Geboes K, Riddell R, Ost A, Jensfelt B, Persson T, Löfberg R. A reproducible grading scale for histological assessment of inflammation in ulcerative colitis. *Gut* 2000;47:404-409.
 10. Stidham RW, Liu W, Bishu S, et al. Performance of a deep learning model vs human reviewers in grading endoscopic disease severity of patients with ulcerative colitis. *JAMA Netw Open* 2019;2:e193963.
 11. Yao H, Najarian K, Gryak J, et al. Fully automated endoscopic disease activity assessment in ulcerative colitis. *Gastrointest Endosc* 2021;93:728-736.
 12. Bossuyt P, Vermeire S, Bisschops R. Scoring endoscopic disease activity in IBD: artificial intelligence sees more and better than we do. *Gut* 2020;69:788-789.
 13. Marchal-Bressenot A, Salleron J, Boulagnon-Rombi C, et al. Development and validation of the Nancy histological index for UC. *Gut* 2017;66:43-49.
 14. Travis SP, Schnell D, Krzeski P, et al. Reliability and initial validation of the ulcerative colitis endoscopic index of severity. *Gastroenterology* 2013;145:987-995.
 15. Takenaka K, Ohtsuka K, Fujii T, et al. Development and validation of a deep neural network for accurate evaluation of endoscopic images from patients with ulcerative colitis. *Gastroenterology* 2020;158:2150-2157.
 16. Takenaka K, Ohtsuka K, Fujii T, Oshima S, Okamoto R, Watanabe M. Deep neural network accurately predicts prognosis of ulcerative colitis using endoscopic images. *Gastroenterology* 2021;160:2175-2177.
 17. Gottlieb K, Requa J, Karnes W, et al. Central reading of ulcerative colitis clinical trial videos using neural networks. *Gastroenterology* 2021;160:710-719.
 18. Takenaka K, Ohtsuka K, Kitazume Y, et al. Comparison of magnetic resonance and balloon enteroscopic examination of the small intestine in patients with Crohn's disease. *Gastroenterology* 2014;147:334-342.
 19. Enns RA, Hookey L, Armstrong D, et al. Clinical practice guidelines for the use of video capsule endoscopy. *Gastroenterology* 2017;152:497-514.
 20. Sturm A, Maaser C, Calabrese E, et al. ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 2: IBD scores and general principles and technical aspects. *J Crohns Colitis* 2019;13:273-284.
 21. Ben-Horin S, Lahat A, Amitai MM, et al. Assessment of small bowel mucosal healing by video capsule endoscopy for the prediction of short-term and long-term risk of Crohn's disease flare: a prospective cohort study. *Lancet Gastroenterol Hepatol* 2019;4:519-528.
 22. Klang E, Barash Y, Margalit RY, et al. Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy. *Gastrointest Endosc* 2020;91:606-613.
 23. Klang E, Grinman A, Soffer S, et al. Automated detection of Crohn's disease intestinal strictures on capsule endoscopy images using deep neural networks. *J Crohns Colitis* 2021;15:749-756.
 24. Barash Y, Azaria L, Soffer S, et al. Ulcer severity grading in video capsule images of patients with Crohn's disease: an ordinal neural network solution. *Gastrointest Endosc* 2021;93:187-192.
 25. Ding Z, Shi H, Zhang H, et al. Gastroenterologist-level identification of small-bowel diseases and normal variants by capsule endoscopy using a deep-learning model. *Gastroenterology* 2019;157:1044-1054.
 26. Colombel JF, Rutgeerts P, Reinisch W, et al. Early mucosal healing with infliximab is associated with improved long-term clinical outcomes in ulcerative colitis. *Gastroenterology* 2011;141:1194-1201.
 27. Hébuterne X, Lémann M, Bouhnik Y, et al. Endoscopic improvement of mucosal lesions in patients with moderate to severe ileocolonic Crohn's disease following treatment with certolizumab pegol. *Gut* 2013;62:201-208.