Journal of the Canadian Association of Gastroenterology, 2020, 3(5), 216–221
doi: 10.1093/jcag/gwz022
Original Article
Advance Access publication 19 June 2019

Original Article

Do You See What I See? An Assessment of Endoscopic Lesions Recognition and Description by Gastroenterology Trainees and Staff Physicians

Lara Hart, MD, FRCPC[1], Mallory Chavannes, MD, MSc, FRCPC[2], Peter L. Lakatos, MD, PhD[1,3], Waqqas Afif, MD, FRCPC[1], Alain Bitton, MD, FRCPC[1], Brian Bressler, MD, MSc, FRCPC[2], Talat Bessissow, MD, FRCPC[1]

[1]Division of Gastroenterology, McGill University Health Center, Montreal, Quebec, Canada; [2]Division of Gastroenterology, University of British Columbia, Vancouver, British Columbia, Canada; [3]Division of Gastroenterology, Semmelweis University, Budapest, Hungary

Correspondence: Lara Hart, MD, FRCPC, McGill University Health Center (MUHC), Montreal General Hospital, 1650 Cedar Ave, Montreal, Quebec H3G 1A4, Canada, e-mail: laramhart@gmail.com

Abstract

Background: Gastroenterologists should accurately describe endoscopic findings and integrate them into management plans. We aimed to determine if trainees and staff are describing inflammatory bowel disease (IBD) lesions in a similar manner.

Methods: Using 20 ileocolonoscopy images, participants described IBD inflammatory burden based on physician severity rating, and Mayo endoscopic score (MES) (ulcerative colitis [UC]) or simple endoscopic score (SES-CD) (Crohn's disease [CD]). Images were selected based on agreement by three IBD experts. Findings of varying severity were presented; 10 images included a question about management. We examined inter-observer agreement among trainees and staff, compared trainees to staff, and determined accuracy of response comparing both groups to IBD experts.

Results: One hundred and twenty-nine staff and 47 trainees participated from across Canada. There was moderate inter-rater agreement using physician severity rating ($\kappa = 0.53$ UC and 0.52 CD for staff, $\kappa = 0.51$ UC and 0.43 CD for trainees). There was moderate inter-rater agreement for MES for staff and trainees ($\kappa = 0.49$ and 0.48, respectively), but fair agreement for SES-CD ($\kappa = 0.37$ and 0.32, respectively). For accuracy of response, the mean score was 68.7% for staff and 63.7% for trainees ($P = 0.028$). Both groups identified healed bowel or severe disease better than mild/moderate ($P < 0.05$). There was high accuracy for management, but staff scored higher than trainees for UC ($P < 0.01$).

Conclusion: Inter-rater agreement on description of IBD lesions was moderate at best. Staff and trainees more accurately describe healed and severe disease, and better describe lesions in UC than CD.

Keywords: *Crohn's disease; Endoscopy; Ulcerative colitis*

Introduction

A core component of gastroenterology training is competency in endoscopy, as this is a fundamental tool used in the diagnosis and management of many diseases (1,2). However, the skills in endoscopy go beyond technical competence. Trainees should be capable of accurately describing findings (1), as this can have significant impact on treatment options. During fellowship, trainees will have diverse exposure to different lesions and may use varied terminology to describe them. Once in practice, staff maintain more consistent terminology in their reports (1,3), however, this varies between individuals and institutions. To date, while there are numerous tools to assess and improve technical competence, there have been no large studies assessing trainees' competence at describing endoscopic lesions. Further, there are no large studies assessing how staff are describing lesions on ileocolonoscopy.

As inflammatory bowel disease (IBD) is one of the most common conditions managed by gastroenterologists, and since endoscopic findings can have a significant impact on patient management, it is imperative that gastroenterologists accurately identify and grade disease severity (4). In particular, the finding of mucosal healing is associated with decrease in need for surgery (5), while active disease is associated with higher rates of disease-related complications and more frequent hospitalizations (4). Both ulcerative colitis and Crohn's disease are associated with specific sets of lesion descriptors (6), however, to allow for ease and consistency in reporting, numerous endoscopic scoring systems were developed (7,8). The routine use of scoring systems has been incorporated into clinical trials to prevent bias and error (9) and they have been found to have high inter-rater agreement between central readers. In a clinical trial by Feagan et al. (2013), using various endoscopic scores for ulcerative colitis, central readers' inter-rater agreement was high with an intraclass correlation (ICC) = 0.78 to 0.83 (10). Similarly, in two clinical trials for Crohn's disease, the Simple Endoscopic Score had high inter-rater agreement by central readers with ICC = 0.77 to 0.86 in one and ICC = 0.83 in the second (11,12). However, it is unknown how scoring systems are being used in clinical practice. Further, it is unclear how scoring systems impact one's assessment of disease severity and need for management modification.

Based on other small studies that have compared experts to nonexperts (13,14), we hypothesize that trainees do not recognize and describe IBD lesions the same way as staff. Further, we suspect that there is a lack of consistency among staff as well. With the objective of improving standardization of reporting for IBD, we conducted a cross-country study to determine if trainees and staff are describing lesions and lesion severity in a similar manner, and if they are using available scoring systems to maintain consistency. Further, we aimed to assess how their reporting impacted choice of management strategy.

METHODS

Study Participants

This cross-sectional questionnaire-based study recruited gastroenterology trainees and staff gastroenterologists from across Canada (March to October 2017). Inclusion criteria were for participants to be (a) gastroenterology trainees (paediatric and adult) or (b) staff gastroenterologists (community and academic centers; paediatric and adult centers). Eligible participants were identified through the Canadian Association of Gastroenterology (CAG) registry, which approved the study, and the questionnaire was e-mailed to the members of CAG via their monthly newsletter. E-mails were sent to division chiefs of academic centers to relay to their staff and trainees. Completion of the questionnaire acted as consent.

Study Design

Participants completed a 20-question image-based questionnaire. Each question presented a patient case (detailing duration of disease, current symptoms, and current management) and an image of a typical finding seen on ileocolonoscopy for UC or CD (9 images UC, 11 images CD). Images varied in severity, with representations ranging from healed mucosa to severe disease (UC: two healed, one mild, three moderate, three severe; CD: two healed, three mild, three moderate, three severe). Each image represented a single segment of bowel, and participants were asked to (a) score the lesion using Mayo Endoscopic Score (MES) for UC, or the Simple Endoscopic Score for Crohn's Disease (SES-CD) (the most widely used scoring systems (6)) for that bowel segment, (b) to describe the severity of the lesion by applying a physician severity rating (PSR): healed, mild, moderate, severe (regardless of their answers for (a)). For half of the questions, participants were also asked to choose a management plan, based on the bowel segment image. For those questions, the options of answers were (a) continue current therapy, (b) escalate therapy, (c) give an induction agent, or (d) de-escalate therapy.

The images were taken from patients diagnosed with IBD (confirmed on clinical, and endoscopic criteria) from 2014 to 2016 (images deidentified). High-quality pictures were selected from *endoworks* © from the ileocolonoscopies of two gastroenterologists (from the McGill IBD center), based on their representation of the variety of lesions seen in IBD (and representing the range of disease severity). The questionnaire has been validated among three IBD experts for internal validity and agreement (two adult gastroenterologists from McGill University, not the same gastroenterologists who contributed the images; one from University of British Columbia) who were blinded to the clinical data and to the others' answers. Twenty-seven questions were provided to the IBD experts—questions with significant variability in response were removed (leading to the final 20 questions used). The questionnaire was built using Google Forms ©. The study was approved by the McGill University Health Center REB (REB 2017–3174).

Objectives

The primary objective was to assess the inter-observer reliability and correlation among the trainees and among the staff. The secondary objective was to assess the variability and accuracy comparing trainees and staff to the IBD experts. We also assessed accuracy of the answers based on disease severity as portrayed by the images.

Statistical Analysis

Categorical data were summarized using frequencies and percentages while continuous variables were summarized using means and standard deviations. An accuracy score was attributed to the participant by giving a score of 1 for each of their answers which matched the expert consensus. Difference in scores between subgroups was assessed using two-sample *t*-test. Differences between the mean agreement with experts for severity level was estimated using one-way analysis of variance. Inter-observer agreement among trainees and staff was calculated using Fleiss Kappa. A Fleiss Kappa result above 0.81 represents very good agreement, between 0.61 and 0.8 represents good agreement, between 0.41 and 0.6 being moderate agreement, and between 0.21 and 0.40 being fair agreement. Sample size was evaluated by looking at previous reports. Daperno et al. reported a kappa of 0.57 (95% confidence interval [CI] 0.51 to 0.62) for MES in 64 physicians who evaluated 6 videos before a training course (15). Based on the standard error formula of kappa in Fleiss and assuming the overall proportion of ratings in each MES category in the current study will be similar to that of Daperno et al., a minimum of 40 raters (each evaluating 10 images) are required to achieve a half width of no more than 0.05 for the 95% CI of kappa (16).

RESULTS

Participant Characteristics

Six hundred and fifty-three members of CAG received the newsletter (including staff and trainees). Given the number of images provided to each participant, power was reached after 40 questionnaires were completed. A total of 176 (27%) physicians from across Canada participated in the study, including 30% of the current trainees: 129 staff and 47 trainees. 116 (89.9%) staff and 35 (74.5%) trainees were adult gastroenterologists. Eighty-six (66.7%) staff worked with trainees on a regular basis. Further characteristics are given in Table 1. Reasons for use or nonuse of the scoring systems are given in Table 2.

Inter-rater Agreement for Lesion Assessment

Looking at the questions for UC, there was moderate inter-rater agreement using PSR for both the staff and trainees (κ = 0.53; 95% CI 0.42 to 0.67 for staff and κ = 0.51; 95% CI 0.41 to 0.63 for trainees). In CD, there was moderate inter-rater agreement for PSR, though trainees had lower agreement here than for UC (κ = 0.52; 95% CI 0.38 to 0.68 for staff and κ = 0.43; 95% CI 0.30 to 0.59 for trainees). The MES inter-rater agreement for both staff and trainees was moderate with κ = 0.49 (95% CI 0.39 to 0.63) for staff and 0.48 (95% CI 0.40 to 0.62) for trainees. In CD, the inter-rater agreement for SES-CD was only fair: κ = 0.37 (95% CI 0.26 to 0.55) for staff and 0.32 (95% CI 0.19 to 0.54) for trainees. Inter-rater agreement is summarized in Table 3.

Table 1. Participants characteristics

	Staff physicians *n* = 129	Trainees *n* = 47
Physicians treating adults, *n* (%)	116 (89.9)	35 (74.5)
Age under 35 years, *n* (%)	18 (14.0)	39 (83.0)
More than 500 colonoscopies performed per year, *n* (%)	76 (58.9)	13 (21.0)
Staff, Time in Practice, *n* (%)		
<10 years	63 (48.8)	
11–20 years	26 (20.2)	
21–30 years	22 (17.1)	
31–40 years	13 (10.1)	
>40 years	3 (2.3)	
Trainees, year in gastroenterology, *n* (%)		
First year		11 (23.4)
Second year		21 (44.7)
Third year		7 (14.9)
Fourth year		3 (6.4)
Fellow		4 (8.5)
Practice environment, *n* (%)		
Academic	78 (60.5)	
Community center – City	44 (34.1)	
Community center – Rural	5 (3.9)	
Use Mayo Endoscopic Score in practice, *n* (%)	114 (88.4)	42 (89.4)
Use SES-CD in practice, *n* (%)	41 (31.8)	26 (55.3)

SES-CD, Simple endoscopic score for Crohn's disease.

Accuracy of Responses (agreement with the experts)

For agreement with experts, the mean overall test score (combining responses from both the UC and CD images) was significantly higher for staff than trainees, with a score of 68.7% (standard deviation [SD] = 13.4%) and 63.7% (SD = 12.2%), respectively (P = 0.028). For staff, there was a statistically higher accuracy for images illustrating UC lesions than CD lesions (72.2% ± 13.2% versus 65.4% ± 17.1%, $P < 0.001$). The results were similar for trainees who had an accuracy score of 66.8 ± 15.5% for UC lesions compared to 60.9 ± 15.7% for CD lesions, though that difference was not statistically significant (P = 0.068). In UC lesions, the PSR had the highest agreement score with experts for both staff and trainees (73.7% ± 16.5% and 70.4% ± 17.8%, respectively) compared to the MES (66.7 ± 15.3% and 62.6 ± 18.6% for staff and trainees, respectively). In CD lesions, we observed a similar trend where PSR had higher agreement with experts for both staff and trainees

Table 2. Reasons why endoscopists use scoring system 50% of the time or less

	Staff physicians (n = 129)	Trainees (n = 47)
Would rather describe the lesions, n (%)	37 (29.7)	13 (27.7)
Forget to use the scoring systems, n (%)	11 (8.5)	2 (4.3)
Not familiar with endoscopic scores, n (%)	10 (7.8)	4 (8.5)
Too complicated to use, n (%)	9 (7.0)	4 (8.5)
Time consuming, n (%)	2 (1.6)	3 (6.4)
Not trained or not used in home center, n (%)	1 (0.8)	3 (6.4)

Table 3. Inter-rater agreement using Fleiss Kappa (κ) for scoring systems

	Staff physicians (n = 129)	Trainees (n = 47)
Mayo Endoscopic Score, κ (95% CI)	0.49 (0.39–0.64)	0.48 (0.40–0.62)
SES-CD, κ (95% CI)	0.37 (0.26–0.58)	0.32 (0.2–0.54)
PSR for UC, κ (95% CI)	0.53 (0.43–0.68)	0.51 (0.41–0.64)
PSR for CD, κ (95% CI)	0.52 (0.39–0.68)	0.43 (0.31–0.59)
Management decision in UC, κ (95% CI)	0.59 (0.49–0.75)	0.41 (0.39–0.63)
Management decision in CD, κ (95% CI)	0.65 (0.71–0.85)	0.56 (0.64–0.78)

CI, Confidence interval; CD, Crohn's disease; PSR, Physician severity rating; UC, Ulcerative colitis; SES-CD, Simple endoscopic score for Crohn's disease.

(68.1 ± 19.3% and 63.8 ± 17.6%, respectively) compared to the SES-CD (55.1 ± 19.6% and 49.7 ± 17.1%, respectively). There was higher agreement with experts on questions representing healed mucosa and severe disease (>75% accuracy for both) compared to mild to moderate disease (<65% accuracy, $P < 0.05$ for both staff and trainees).

Choice of Management Strategy (based on image appearance)

There was moderate agreement on management plan for UC ($\kappa = 0.59$, 95% CI = 0.49 to 0.76), and good agreement in CD ($\kappa = 0.65$, 95% CI = 0.71 to 0.85) for staff. Trainees had moderate inter-rater agreement for both of UC and CD management questions ($\kappa = 0.41$, 95% CI = 0.37 to 0.63 and $\kappa = 0.56$, 95% CI = 0.64 to 0.78, respectively). The accuracy of response for management plan was high for both staff and trainees (Staff: UC = 78.0 ± 20.9%, CD = 86.6 ± 24.0%, Trainees: UC = 67.7 ± 19.2%, CD = 83.5 ± 24.0%). Staff had significantly higher agreement with experts than trainees for the management plan in UC ($P = 0.004$) but not for CD ($P = 0.45$).

Discussion

In this real-world study of endoscopic lesion recognition, there was moderate inter-rater agreement for MES (UC), fair agreement for SES-CD (CD) and moderate inter-rater agreement for PSR (UC + CD) for both staff and trainees. Both groups identified healed bowel or severe disease better than mild/moderate, and there was high accuracy of response for management strategies (compared to experts), with staff scoring higher than trainees for UC but not CD.

From the questionnaire, 88% of staff and 89% of trainees describe using the MES in practice, but only 32% of staff and 55% of trainees use the SES-CD. In the literature, the SES-CD is noted to be difficult to use, given its complexity and numerous subcomponents that need to be calculated (6). On the other hand, the strength of the MES is its ease of use (6), which could explain why scores for UC were more accurate for both staff and trainees. However, it remains puzzling that trainees are using the SES-CD at a higher rate than their educators. Furthermore, participants cited their top reason for not using scoring systems was 'would rather describe the lesions', while less than 10% said the scoring systems were too complicated and time consuming.

Without any formal training in scoring systems, the staff and trainees' inter-rater agreement in our study was not nearly similar to the clinical trial findings of trained central readers (with an ICC 0.78 to 0.83 for UC using various scoring systems (10) and 0.77 to 0.86 for SES-CD (11,12)). However, in comparison to other studies that included untrained reading of endoscopy, staff in our study had similar inter-rater agreement for MES and SES-CD. In these studies, the ICC for MES ranged from 0.46 to 0.57 (13,17,18), while for SES-CD, the ICC ranged from 0.69 to 0.78 (10,11,15,19). It is unclear if this is due to presentation of images (rather than videos) or a different countrywide and much larger cohort of participants in our study. Nevertheless, it is now evident that gastroenterologists should rely on endoscopic disease severity more than clinical symptoms, as the latter do not correlate with endoscopic findings (2). Therefore, based on these real-world findings, further training in the use of scoring systems may be warranted, especially since their use could offer uniformity in endoscopy reporting in day-to-day practice, and in clinical decision making. Using scoring systems in reports could also be useful in trending patients' progress or when transferring care to another physician.

Unique to our study, we were able to determine how well a large group of gastroenterologists identified disease severity. While overall accuracy of response for all lesions was less than 70%, both trainees and staff had strong acumen and judgement (high accuracy) in their ability to identify both healed bowel and severe disease (but less so for mild-moderate disease). In two previous studies, gastroenterologists had high inter-rater agreement for healed bowel segments (20,21). Orlandi et al (1998) also found high inter-observer agreement of 15 gastroenterologists for severe disease, similar to our study. With the presence of multiple biologic agents available for clinical use for IBD, it is important that gastroenterologists recognize drug effectiveness as defined by the presence of healed bowel on repeat colonoscopy (21,22). Conversely, it is important that gastroenterologists are able to identify severely active disease and modify management to decrease the risk of complications (15).

This is the first large study to assess trainees' ability to score IBD lesions on ileocolonoscopy. A small study by Osada (2010) with four trainees assessing UC lesions found an inter-rater agreement for MES of 0.44 to 0.47, similar to our findings (13). Overall, inter-rater agreement for trainees was fair (SES-CD) to moderate (for MES, UC PSR and CD PSR), which was surprisingly similar to the findings for staff. However, when assessing accuracy of response compared to IBD experts, staff scored significantly higher than trainees (for both UC and CD). Given their breadth of experience, this latter finding was not unexpected. Hyun et al. (2013) also found similar results when assessing the endoscopic diagnostic accuracy of gastric metaplasia between staff and trainees (23).

This is the only IBD study to assess both identification of IBD lesions and associated management plan (based on the visual findings). Based on best practice, if active disease is identified on endoscopy, therapy should be modified accordingly. Our study provides evidence that gastroenterologists are applying this principle, as staff had a high accuracy of response for management when compared to the IBD experts (regardless of discrepancies in scoring). Furthermore, there was a higher accuracy in response for physician severity rating than with either of MES or SES-CD. It could therefore be postulated that as gastroenterologists gain experience, they are better able to qualify images as healed, mild, moderate and severe and use this grading to guide therapeutic changes.

There were limitations to this study. Sixty per cent of staff responders were from academic centers, which are not representative of the practice landscape and therefore may impact generalizability. However, all the trainees receive their education from academic centers; therefore, it is helpful to compare trainees to academic physicians. Response rate was 27% of gastroenterologists and trainees across Canada, despite reminders. However, given the number of questions posed to each participant, the study was well powered, as power was reached after 40 participants completed the questionnaire. Only two of the available scoring systems were utilized in our questionnaire. Therefore, it is possible that if other scoring systems had been used, there may have been higher inter-rater agreement and accuracy of response. We chose the MES and SES-CD since they are most widely used in clinical trials, are most easy to use and are recommended for day-to-day practice (6,8,21). The questionnaire used still images of sections of the bowel rather than videos. In particular, this may have affected one's ability to accurately use SES-CD, given that it asks for percent affected area in a bowel segment (8). While videos are more representative of what is seen during colonoscopy, it would have made the questionnaire longer. We suspect this would have led to a decreased response rate and/or would have necessitated providing fewer images, thereby potentially compromising the power of the study. As gastroenterologists take pictures during the colonoscopy, we believed the still images represented what would be seen in a typical endoscopy report (and the participants could apply SES-CD and MES to what was visually presented in the image). The questionnaire did require at least 15 to 30 minutes to complete. This could have led to rushing through the final few questions, which could have affected the responses. However, there is no process to verify if this occurred. Moreover, the number of questions/subquestions in the study allowed us to assess identification and classification of lesions of different severity and associated management strategy, which has not been done before.

CONCLUSION

In this large study, inter-rater agreement on description of ileocolonic lesions in IBD is moderate at best. Both staff and trainees more accurately describe lesions in UC than in CD and are more accurate using severity rating (healed, mild, moderate, severe) than the endoscopic scoring systems. Healed bowel or severe disease is more accurately described than mild/moderate disease. Further efforts are needed to identify the optimal means of standardizing reporting of IBD lesions in a clinical practice setting, with instituting formal teaching during gastroenterology training as one consideration for how to do this.

Funding

The funding for this study was provided by the Canadian Association of Gastroenterology Education Research Grant 2017. The funding source had no role in the design, practice or analysis of this study.

Author Contributions

L.H. and M.C. conceptualized and designed the study, organized the data for analysis, analyzed the data, drafted the initial manuscript and approved the final manuscript as submitted. W.A., A.B., P.L.L., and B.B. aided in interpretation of data, ensured data integrity, revised the manuscript and approved the final manuscript as submitted. T.B. is the supervising author. He assisted in the conceptualization and design of the trial, monitored data collection, critically reviewed and revised the manuscript and approved the final manuscript as submitted.

References

1. Walsh CM. In-training gastrointestinal endoscopy competency assessment tools: Types of tools, validation and impact. Best Pract Res Clin Gastroenterol 2016;30(3):357–74.
2. Neurath MF, Travis SP. Mucosal healing in inflammatory bowel diseases: A systematic review. Gut 2012;61(11):1619–35.
3. Devlin SM, Melmed GY, Irving PM, et al. Recommendations for quality colonoscopy reporting for patients with inflammatory bowel disease: Results from a RAND appropriateness panel. Inflamm Bowel Dis 2016;22(6):1418–24.
4. 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(4):1194–201.
5. Reinink AR, Lee TC, Higgins PD. Endoscopic mucosal healing predicts favorable clinical outcomes in inflammatory bowel disease: A meta-analysis. Inflamm Bowel Dis 2016;22(8):1859–69.
6. Christensen B, Rubin DT. Understanding endoscopic disease activity in IBD: How to incorporate it into practice. Curr Gastroenterol Rep 2016;18(1):5.
7. Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. N Engl J Med 1987;317(26):1625–9.
8. Daperno M, D'Haens G, Van Assche G, et al. Development and validation of a new, simplified endoscopic activity score for Crohn's disease: The SES-CD. Gastrointest Endosc 2004;60(4):505–12.
9. Panés J, Feagan BG, Hussain F, et al. Central endoscopy reading in inflammatory bowel diseases. J Crohns Colitis 2016;10(Suppl 2):S542–7.
10. Feagan BG, Sandborn WJ, D'Haens G, et al. The role of centralized reading of endoscopy in a randomized controlled trial of mesalamine for ulcerative colitis. Gastroenterology 2013;145(1):149–157.e2.
11. Rutgeerts P, Reinisch W, Colombel JF, et al. Agreement of site and central readings of ileocolonoscopic scores in Crohn's disease: Comparison using data from the EXTEND trial. Gastrointest Endosc 2016;83(1):188–97.e1–3.
12. Khanna R, Zou G, D'Haens G, et al. Reliability among central readers in the evaluation of endoscopic findings from patients with Crohn's disease. Gut 2016;65(7):1119–25.
13. 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(2):192–7.
14. de Lange T, Larsen S, Aabakken L. Inter-observer agreement in the assessment of endoscopic findings in ulcerative colitis. BMC Gastroenterol 2004;4:9.
15. Daperno M, Comberlato M, Bossa F, et al. Inter-observer agreement in endoscopic scoring systems: Preliminary report of an ongoing study from the Italian Group for Inflammatory Bowel Disease (IG-IBD). Dig Liver Dis 2014;46(11):969–73.
16. Fleiss JL, Levin B, Paik MC. Statistical Methods for Rates and Proportions. Hoboken, NJ: John Wiley & Sons, Inc, 2003.
17. Daperno M, Comberlato M, Bossa F, et al. Training programs on endoscopic scoring systems for inflammatory bowel disease lead to a significant increase in interobserver agreement among community gastroenterologists. J Crohns Colitis. 2016;8:jjw181.
18. Fernandes SR, Pinto JSLD, Marques da Costa P, et al.; GEDII. Disagreement among gastroenterologists using the mayo and rutgeerts endoscopic scores. Inflamm Bowel Dis 2018;24(2):254–60.
19. Dubcenco E, Zou G, Stitt L, et al. Effect of standardised scoring conventions on inter-rater reliability in the endoscopic evaluation of Crohn's disease. J Crohns Colitis 2016;10(9):1006–14.
20. Orlandi F, Brunelli E, Feliciangeli G, et al. Observer agreement in endoscopic assessment of ulcerative colitis. Ital J Gastroenterol Hepatol 1998;30(5):539–41.
21. Walsh A, Palmer R, Travis S. Mucosal healing as a target of therapy for colonic inflammatory bowel disease and methods to score disease activity. Gastrointest Endosc Clin N Am 2014;24(3):367–78.
22. Levesque BG, Sandborn WJ, Ruel J, et al. Converging goals of treatment of inflammatory bowel disease from clinical trials and practice. Gastroenterology 2015;148(1):37–51.e1.
23. Hyun YS, Han DS, Bae JH, et al. Interobserver variability and accuracy of high-definition endoscopic diagnosis for gastric intestinal metaplasia among experienced and inexperienced endoscopists. J Korean Med Sci 2013;28(5):744–9.