Scoring Endoscopy in Pediatric Inflammatory Bowel Disease: A Way to Improve Quality

*Lorenzo Norsa, †Alberto Ferrari, ‡Serena Arrigo, §Matteo Bramuzzo, ||Marco Deganello Saccomani, ¶Giovanni Di Nardo, #Maria Teresa Illiceto, **Erasmo Miele, ††Monica Paci, ‡‡Claudio Romano, §§Erminia Romeo, ||||Marco Daperno, and ¶¶Salvatore Oliva, on behalf of the Endoscopy group of the Italian society of pediatric Gastroenterology Hepatology and Nutrition (SIGENP)

ABSTRACT

Objectives and study: There is a large interobserver variability in evaluating mucosal lesions of inflammatory bowel disease (IBD), especially in pediatric patients. This multicenter prospective observational study aims to evaluate interobserver agreement (IOA) among pediatric endoscopists in assigning validated IBD endoscopic scores in children.

Methods: Fifteen videos of follow-up ileocolonoscopies in children with IBD were recorded and selected as cases. Eleven pediatric endoscopists from different centers blindly evaluated all videos and calculated scores: either Ulcerative Colitis Endoscopic Index of Severity (UCEIS) or Simple Endoscopic Score for Crohn Disease (SES-CD). Scores from all reviewers were compared in order to calculate IOA for general videos and specific sections. Scores from an expert adult reader were used to calculate possible reviewer's characteristics affecting scores' reliability.

Results: Intraclass correlation was 0.298 (95% confidence interval [CI]: 0.13–0.55) for ulcerative colitis (UC) and 0.266 (0.11–0.52) for Crohn disease (CD). When a disease activity categorization was adopted (remission, mild, moderate, severe activity) Fleiss kappa coefficient was 0.408 (0.29–0.53) for UC and 0.552 (0.43–0.73) for CD. When stratified by item, vascular pattern of UC was the most reliable item IC: 0.624 (0.321–0.854). In multivariable analysis, none of the reviewer's characteristics affected the readers' errors.

Conclusions: This multicenter study shows low agreement among pediatric endoscopists in evaluating endoscopic scores in children with IBD. By using disease activity categorization, agreement slightly increased, mostly for CD. All readers showed a low-grade concordance with the expert adult gastroenterologist's evaluations. Future-specific training programs should be considered to increase IOA in using IBD endoscopic activity scores.

Key Words: Crohn disease, endoscopic score, ulcerative colitis

(*JPGN* 2021;72: 00–00)

What Is Known

- Mucosal healing is the ultimate goal of pediatric inflammatory bowel disease treatment.
- Endoscopy during follow-up is crucial to assess mucosal healing in pediatric inflammatory bowel disease.
- Endoscopic scoring system are available to grade mucosal inflammation in inflammatory bowel disease patients.

What Is New

- Pediatric endoscopists show a poor grade of agreement in scoring endoscopies from pediatric inflammatory bowel disease.
- When stratified for disease activity, agreement slightly increases especially for Crohn disease.
- No endoscopists' related factors are linked to the low grade of reproducibility.

The advent of mucosal healing (MH) as the ultimate therapeutic goal in pediatric inflammatory bowel disease (IBD) (1) has given endoscopy a central role in the management of IBD. MH has been considered as the main treatment target in clinical trials (2), and is now widely used to guide treatments in a treat-to-target strategy, both in adults and children (3).

Received November 11, 2020; accepted February 6, 2021.

From the *Pediatric Gastroenterology Hepatology and Transplantation, the †FROM Research Foundation, ASST Papa Giovanni XXIII, Bergamo, the ‡Pediatric Gastroenterology and Endoscopy Unit, G. Gaslini Institute-IRCCS, Genova, the §Gastroenterology, Digestive Endoscopy and Nutrition Unit, Institute of Child and Maternal Health-IRCCS ''Burlo Garofolo'', Trieste, the ||Pediatric Department, Women's & Children's Hospital, AOUI of Verona, the ¶Pediatrics, NESMOS Department, School of Medicine and Psychology, Sapienza University of Rome, Sant'Andrea University Hospital, Rome, the #Pediatric Gastroenterology and Endoscopy Unit, ''S. Spirito'' Hospital, Pescara, the **Department of Translational Medical Science, Section of Pediatrics, University of Naples ''Federico II'', Napoli, the ††Pediatric Gastroenterology and Endoscopy Unit, ''Mayer'' Hospital, Florence, the ‡‡Pediatric Gastroenterology and Cystic Fibrosis Unit, University of Messina, the §§Digestive Endoscopy and Surgery Unit, Bambino Gesù Children Hospital, IRCCS, Rome, the ||||Gastroenterology Unit, Mauriziano Hospital, Turin, and the ¶¶Pediatric Gastroenterology and Liver Unit, Maternal and Child Department, Sapienza – University of Rome, Italy.

Address correspondence and reprint requests to Lorenzo Norsa, Pediatric Hepatology Gastroenterology and Transplantation, ASST Papa Giovanni XXIII, Piazza OMS1, 24027 Bergamo, Italy (e-mail: lnorsa@asst-pg23.it

Supplemental digital content is available for this article. Direct URL citations appear in the printed text, and links to the digital files are provided in the HTML text of this article on the journal's Web site (*www.jpgn.org*).

The authors report no conflicts of interest.

DOI: 10.1097/MPG.0000000000003090

The European Society of Pediatric Gastroenterology Hepatology and Nutrition (ESPGHAN) position paper strongly encourages clinicians taking care of pediatric IBD to adopt endoscopy activity scores not only in clinical trials but also more generally in clinical practice (4). It also addresses the use of specific endoscopic scores to monitor disease activity, pointing out strengths and limitations of each scoring system (5,6).

Several efforts have been made in order to validate and to assess reliability of endoscopic activity scores in adult IBD patients (7–11).

Despite those endoscopic activity indexes being widely used in pediatric clinical trials (12–15), currently no study has assessed the performance of pediatric endoscopists in using adult IBD endoscopy scores in real practice.

The aim of this study was to assess interobserver agreement of pediatric endoscopists in scoring pediatric IBD by using validated endoscopy activity scores.

MATERIAL AND METHODS

An open call within the Endoscopy Working Group of the Italian Society of Pediatric Gastroenterology Hepatology and Nutrition (SIGENP) was made to involve all the tertiary pediatric IBD centers in a project on endoscopic activity scores.

A group of 11 fully trained pediatric endoscopists from 11 different tertiary centers for pediatric IBD agreed to participate. All participants provided information regarding their experience in pediatric endoscopy and the number of IBD children followed in their unit. No specific training was required for this study as it was meant to assess competence in a common clinical practice setting.

A call for recorded videos of complete ileocolonoscopies was made among pediatric IBD centers. Inclusion criteria for video were a full registration of colonoscopy from the terminal ileum until the rectum before biopsy samples. Video needed to be longer than 5 minutes presented in mp4 format with a high-quality registration. Three IBD centers in different geographical areas over the country (Rome, Naples, and Trieste) provided 15 videos after anonymization. Seven videos were taken on Crohn disease (CD) and 8 from ulcerative colitis (UC). Video were provided to readers, not cut. The number of video was decided in order to be comparable with previous adult studies (16,17) and to guarantee the best compliance of the readers.

The only piece of clinical information provided to the participant endoscopists was IBD diagnosis (UC or CD); no data on clinical activity were available. Due to design study, no IRB approval was necessary.

Participants were asked to review and score videos according to their experience and judgment. The following scoring systems were used: the Ulcerative Colitis Endoscopic Index of Severity (UCEIS) (18) and the Simple Endoscopic Score for Crohn Disease (SES-CD) (9) for UC and CD, respectively. UCEIS score was preferred to traditional and simpler Mayo score because of better outcomes described in predicting clinical outcomes and long-term prognosis of ulcerative colitis patients (19–21).

Reviewers provided their scores on an electronic Excel spreadsheet shared on a specific online platform. The scores were presented as total score per video, and separated categories' subscores.

UC items were: vascular pattern, bleeding, erosion, and ulcers; whereas CD items were: size of ulcers, ulcerated surface, affected surface, and presence of stenosis. For CD, each category was separately evaluated for each ileocolic segment (ileum, cecum, traverse, descending colon, and rectum).

Sub-classes were calculated for each score according to the disease severity. For SES-CD ''quiescent activity'' was defined for score <3, ''mild'' between 4 and 10, ''moderate'' between 11 and 17, and ''severe'' for score above 18 (22). For UCEIS, ''quiescent activity'' was defined for score 0, ''mild'' activity <3, moderate between 3 and 5, and severe for score above 7 (19).

Diseases were dealt with separately, and for each disease, a full model including all covariates and random effects was fitted first to identify the possible factors that affect the interobserver agreement (IOA). An adult international expert in this field (M.D.) was invited to revise and score the same 15 videos as external reference. Those scores were used in the analysis to determine possible specific endoscopist's related variable affecting pediatric endoscopists' results.

Statistical Analysis

Calculation of intraclass correlation coefficients was performed by analysis of variance on a series of Generalized Linear Mixed Models (GLMMs): a logistic, a probit, and a classic mixed linear model. CD and UC were treated separately.

For each of the 2 scores, at first, a complete model was fitted accounting for all the variability in the data. These mixed effect models consisted of 4 components:

1. A fixed effects component including as covariates' age, endoscopist's experience, and IBD follow-up, that is, a series of characteristics of the operator.
2. A random effect variance component for ''Operator'': this is a variance parameter accounting for any additional variation that is because of further individual differences between operators.
3. A second random effect variance component for ''Video'' accounting for the additional between-video variation.
4. Residual random error: a random error term.

The model assumes the following form:

$$g(y|X) = \beta' X + \gamma' z + \delta' w + \varepsilon$$

Where g is a link function (logit, probit or identity), the β' is the vector of fixed effect coefficients, X is a matrix of covariates, γ' is the vector of operator-specific random coefficients with variance σ^2_γ, z is the vector of operator indicators, δ' is the vector of video-specific random coefficients with variance σ^2_δ, w is the vector of video indicators, and ε is a vector of random Gaussian errors with variance σ^2_ε. Intraclass correlation coefficient (ICC) are estimated as the variance partition of between-video variance over total variance

$$\frac{\sigma^2_\delta}{\sigma^2_\gamma + \sigma^2_\delta + \sigma^2_\varepsilon}$$

Agreement was also evaluated by discretizing scores into classes (mild, moderate, severe) and applying Fleiss kappa. Following Fleiss own guidelines (23), values of kappa higher than 0.75 were deemed excellent, values constituted between 0.40 and 0.75 fair/good, and below 0.40 low.

RESULTS

The characteristics of the 11 pediatric endoscopists enrolled in the study are presented in Table 1, Supplemental Digital Content *http://links.lww.com/MPG/C248*.

The first analysis calculated the overall IOA. Table 2 summarizes ICCs obtained with different models along with Fleiss kappa values on the categorized scores. ICCs calculated by probity and logit models are very much overlapping and appear quite low, whereas those obtained by a linear model are very high, and Fleiss kappa values are somewhat in between.

AQ7 TABLE 1. Various methods of calculation of interobserver agreement

	Intracluster correlation coefficient				Fleiss kappa			
Crohn disease	Model	ICC	Standard error	95% CI	Activity	Kappa	*z*	95% CI
	Logit	0.279	0.111	0.116–0.532	Quiescent	0.804	15.8	
	Probit	0.267	0.107	0.111–0.514	Mild	−0.026	−0.52	
	Linear	0.829	0.081	0.613–0.937	Moderate	0.603	11.84	
					Severe	0.597	11.71	
					Combined	0.633	18.24	0.531–0.801
Ulcerative colitis	Model	ICC	Standard error	95% CI	Activity	Kappa	*z*	95% CI
	Logit	0.325	0.1164	0.145–0.576	Quiescent	0.694	14.55	
	Probit	0.360	0.1211	0.167–0.612	Mild	0.415	8.71	
	Linear	0.844	0.0701	0.656–0.939	Moderate	0.309	6.47	
					Severe	0.109	2.28	
					Combined	0.404	13.58	0.293–0.529

CI = confidence interval; ICC = intraclass correlation coefficient.

For what concerns CD, it was found that, once the model was adjusted for the significant operator-specific covariates (age, etc), the variance estimate for the Operator random effect is close to zero; the Operator effect is, therefore, dropped from the model out of parsimony.

For what concerns UC, none of the operator-specific covariates is significant, whereas between-operators' variance is different from zero. Therefore, in this case, both fixed effect covariates and random operator effects are dropped, and a fixed operator effect is introduced instead, to account for all the variables because of individual differences between scorers. This choice also allowed us to end up with a model with 1 random effect, which is more computationally treatable.

After classes' stratification, IOA is 0.633 (95% CI: 0.531–0.801) for CD and 0.404 (95% CI: 0.293–0.529) for UC (Table 1).

Separated results for every video are shown in Figure 1.

We also stratified IOA for item by a logistic model; scores are showed in Table 2. The most reproducible evaluation in our

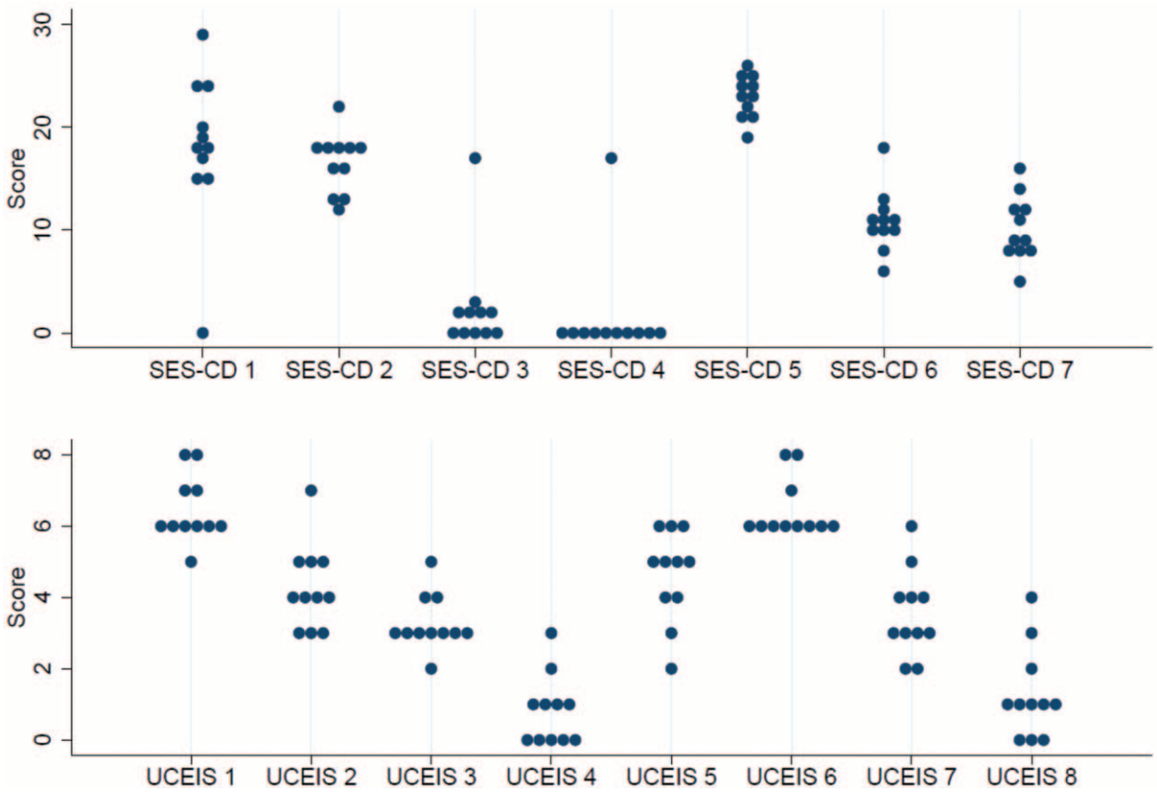

FIGURE 1. Scoring results divided by video ID. SES-CD = Simple Endoscopic Score for Crohn Disease; UCEIS = Ulcerative Colitis Endoscopic Index of Severity.

TABLE 2. Stratification of interobserver agreement according to item scored

Sub-scores	ICC	95% confidence intervals
SES-CD		
Size of ulcers	0.276	0.111–0.539
Ulcerated surface	0.303	0.121–0.577
Affected surface	0.304	0.124–0.576
Stenosis	0.307	0.092–0.660
UCEIS		
Vascular pattern	0.624	0.321–0.854
Bleeding	0.310	0.121–0.594
Erosion and ulcers	0.327	0.133–0.607

ICC = intraclass correlation coefficient; SES-CD = Simple Endoscopic Score for Crohn Disease; UCEIS = Ulcerative Colitis Endoscopic Index of Severity.

cohort of endoscopists is the vascular pattern in UC and the least reproducible is expressed by the size of ulcers in CD.

The multivariate analysis does not find any significant factor affecting readers' evaluation accuracy (Table 2, Supplemental Digital Content, *http://links.lww.com/MPG/C249*).

In the last analysis, all scores are compared with an external reference represented by scores calculated by the expert adult endoscopist. Results of this analysis are shown in Figure 2. Readers display scattered means of errors compared with reference ranging from a minimum of 2.29 to a maximum of 9.14 for SES-CD, and from a minimum of 0.63 to a maximum of 2.13 for UCEIS. When clustered for disease activity, the concordance is between 14% and 85% for SES-CD, and between 37.5% and 75% in UCEIS.

DISCUSSION

This study indicates that a large variability exists between pediatric endoscopists while using endoscopic scoring systems of IBD disease activity. The item, which showed the highest reproducibility was the vascular pattern in UC score. It was not possible to identify any readers' characteristics affecting the overall ability to use IBD scores.

Reproducibility is an essential requirement to provide quality in endoscopy; this is particularly true in the field of IBD, considering that the state of mucosal inflammation might not be easily assessed in an objective manner, and it affects therapeutic decision-making. Despite being widely validated, scoring systems are usually difficult to be calculated on recorded images or videos, as the indirect evaluation extremely increases the variability and subjectivity of operators' assessments. For this reason, clinical trials usually apply a central reading to reduce variability (24,25). Recent discussion has, however, emphasized the role of reliable scoring as a goal for improving clinical practice (26).

Recent studies have evaluated interobserver agreement in IBD endoscopic scores, in particular, one derived from a large clinical trial (25). In this study, IOA for CDEIS and SES-CD was very high, with values above 0.75 for both scores at any time-point (baseline, 12 and 52 weeks). Those values are remarkably high compared with ours but they have been produced in a well-defined clinical trial setting and by involving highly experienced IBD endoscopists. Our study focused on a cohort of pediatric endoscopists in a real practice setting, thus not strictly requiring high levels of expertise. Furthermore, another possible explanation of our lower results is readers of the mentioned trial were aware of the patients' medical history.

An important effort in evaluating IOA in IBD scores in clinical practice was made by the Italian group of study in

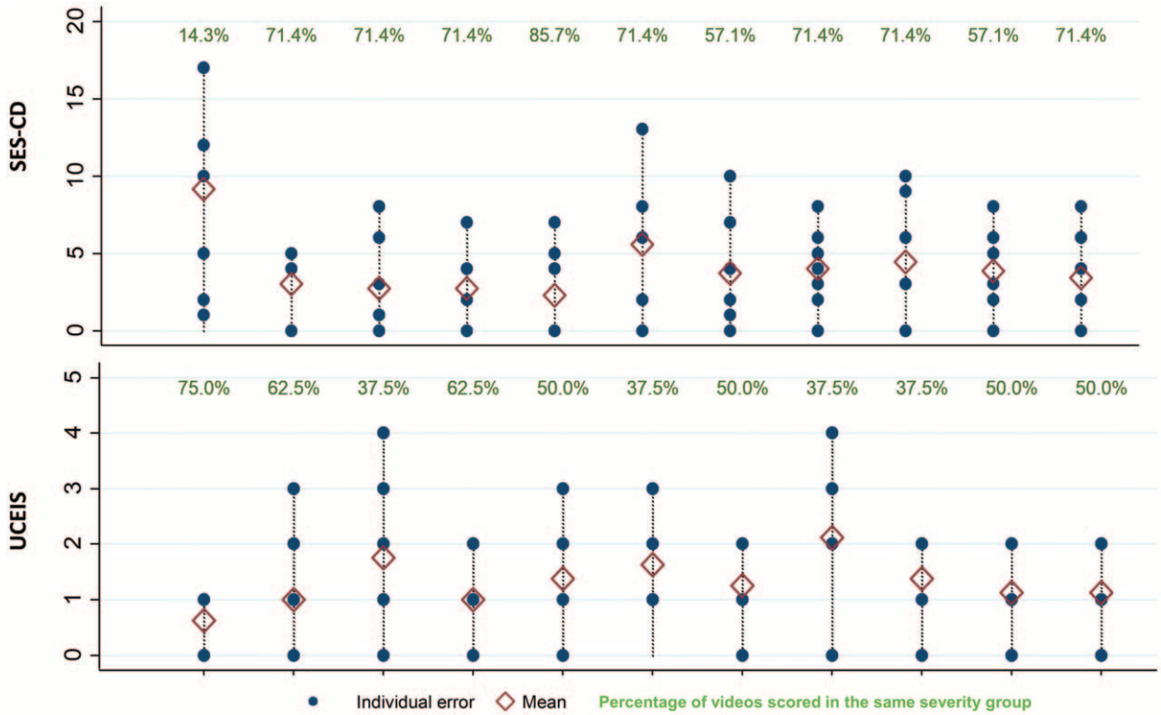

FIGURE 2. Agreement of singular reader with adult gastroenterologist. SES-CD = Simple Endoscopic Score for Crohn Disease; UCEIS = Ulcerative Colitis Endoscopic Index of Severity.

IBD (IG-IBD) (16,17). In an article published in 2014, Mayo score was used to evaluate UC resulting in an ICC of 0.53 and 0.51, for experienced and unexperienced endoscopists in IBD, respectively. For CD cases, they showed an important difference in ICC of SES-CD between experts and nonexperts, with values accounting for 0.93 and 0.68, respectively (16).

The same group published another work focusing on the impact of training on reading ability of endoscopists without specific experience in IBD. Results were encouraging, with an increase in IOA for Mayo score from 0.51 to 0.76, after a training program. Unfortunately, the analysis for CD was only made at baseline, showing an ICC of 0.77 when using SES-CD.

A possible explanation of the difference between our results with those from previous studies on IOA is determined by the method used for ICC calculations. Usually, ICC is estimated from a random effects model as a variance partition of between-objects variance over total variance; however, estimates of both components show considerable variation depending on the specific model used to derive them and its assumptions (27). In the present work, we show that choice of the model can indeed increase/decrease ICC estimates 2- to 3-fold (Table 2); furthermore, ICC poorly reflects agreement in clinical evaluation as evaluated by the kappa statistic, especially when the former is calculated by a linear model. We assume that previous studies on the topic modelled ICC in the general framework of linear modelling, yet, based on the nature of endoscopic scores that are bound in an interval, it seems more sensible to adopt a binomial model with logit or probit link function. Our data suggest that calculation of ICC based on the linear framework may indeed lead to overestimation of the agreement, which would explain discrepancies between the very high ICC obtained by the linear model on the raw scores and the much less impressive values of Fleiss kappa values on the categorized outcome.

Unfortunately, to our knowledge, no comprehensive evaluation of the impact of model choice and data structure on ICC has yet been performed. A simulation study by Wu et al (28) compared different methods for ICC estimation on binary data and showed that its estimate does indeed vary dramatically with estimation methods but their analysis was not focused on random intercept binomial models so much as on linear methods and Generalized Estimating Equations. This makes choice of the appropriate method and interpretation of ICC in this context problematic. It is, however, expected that ICC estimates should be consistent when calculation method is unchanged, which means that, as far as ICC are calculated by the same method, comparisons between them are still meaningful. In particular, in our results for all cases, IOA turns out higher in CD than in UC. For the time being, we warn from excessive confidence in interpreting very high ICC as evidence of very good agreement in scores on a bound scale, until the reliability of these calculations is thoroughly tested.

With all the aforementioned limitations, several considerations could be drawn.

First of all, pediatric endoscopists display absolute ICC values lower than 0.75 and, in some cases, lower than 0.4, leading to a low IOA. Poor outcomes could be caused by lower experience of pediatric endoscopists in classifying endoscopic lesions by validated IBD scores. Adult endoscopists have always faced the difficult task of grading polypoid formations for the screening of colorectal cancer, thus they are more familiar with objective evaluations of lesions. Furthermore many efforts were made to increase IOA in this specific setting (29–31).

Better Fleiss kappa was shown for patients with milder activity for both CD and UC (Table 1). Those results are probably because of a higher exposure of pediatric endoscopists to normal examinations, which makes easier to score minimal lesions (4).

According to our results, it is difficult to identify the most specific reason explaining these lower values among pediatricians. Perhaps, as the current IBD scores have been developed specifically in an adult setting, pediatricians could face problems in describing small and more variable lesions that are frequently identified in pediatric patients. Historically, documentation of endoscopic disease activity in children remains generally problematic and related to the operators' judgment (4). The presence of subtle and discontinuous lesions makes an objective evaluation more challenging and imprecise for both pediatric and adult endoscopists in children.

Results commented could indicate a need of specific training in IBD endoscopic scoring for pediatric endoscopists, to achieve a satisfactory reproducibility. All pediatric endoscopists may undergo such a training, regardless of their experience in pediatric endoscopy or IBD. Such training should be focused on pediatric IBD patients in order to depict specific pattern of lesions presented in children.

Some limitations of this study should be acknowledged. First, the small number of centers involved. A recent survey on Italian pediatric endoscopy, however, showed that only 19 centers specifically perform endoscopy in children in Italy, thus, our cohort of 11 sites represents more than 50% of the total available centers in the country (32). Second, the limited number of analyzed videos could also affect results. It is, however, very unlikely that increasing the number of videos would have changed our outcomes, considering the wide variability presented by participants for each video (Fig. 1), not affected by readers' experience or type of analyzed lesions.

CONCLUSIONS

Pediatric endoscopists show a very low agreement in using validated endoscopic scores to assess IBD mucosal lesions. On the basis of adult gastroenterologists' experience, specific training seems to improve agreement; thus, national and international pediatric gastroenterology societies should make an effort in organizing such a training for pediatric endoscopists. By improving training, it might be possible to evaluate if endoscopic scores, validated for adult IBD, are really applicable in children or if specific scores needs to be implemented in the pediatric clinical practice.

REFERENCES

1. Peyrin-Biroulet L, Sandborn W, Sands BE, et al. Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE): determining therapeutic goals for treat-to-target. *Am J Gastroenterol* 2015;110:1324–38. AQ6
2. Ruemmele FM, Hyams JS, Otley A, et al. Outcome measures for clinical trials in paediatric IBD: an evidence-based, expert-driven practical statement paper of the paediatric ECCO committee. *Gut* 2015;64:438–46.
3. Santha SL, Shankar PR, Pan A, et al. Mucosal healing in clinical practice: a single-center pediatric IBD experience. *Inflamm Bowel Dis* 2017;23:1447–53.
4. Oliva S, Thomson M, de Ridder L, et al. Endoscopy in pediatric inflammatory bowel disease: a position paper on behalf of the Porto IBD Group of the European Society for Pediatric Gastroenterology, Hepatology and Nutrition. *J Pediatr Gastroenterol Nutr* 2018;67:414–30.
5. Khanna R, Nelson SA, Feagan BG, et al. Endoscopic scoring indices for evaluation of disease activity in Crohn's disease. *Cochrane Database Syst Rev* (8):2016:CD010642.
6. Mohammed Vashist N, Samaan M, Mosli MH, et al. Endoscopic scoring indices for evaluation of disease activity in ulcerative colitis. *Cochrane Database Syst Rev* 2018;1:CD011450.
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;317:1625–9.
8. Mary JY, Modigliani R. Development and validation of an endoscopic index of the severity for Crohn's disease: a prospective multicentre study. Groupe d'Etudes Thérapeutiques des Affections Inflammatoires du Tube Digestif (GETAID). *Gut* 1989;30:983–9.

9. 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;60:505–12.
10. Travis SPL, Schnell D, Krzeski P, et al. Reliability and initial validation of the ulcerative colitis endoscopic index of severity. *Gastroenterology* 2013;145:987–95.
11. Ferrante M, Colombel J-F, Sandborn WJ, et al., International Organization for the Study of Inflammatory Bowel Diseases. Validation of endoscopic activity scores in patients with Crohn's disease based on a post hoc analysis of data from SONIC. *Gastroenterology* 2013;145:978.e5–86.e5.
12. Puolanne A-M, Kolho K-L, Alfthan H, et al. Rapid fecal calprotectin test and symptom index in monitoring the disease activity in colonic inflammatory bowel disease. *Dig Dis Sci* 2017;62:3123–30.
13. Ricciuto A, Fish J, Carman N, et al. Symptoms do not correlate with findings from colonoscopy in children with inflammatory bowel disease and primary sclerosing cholangitis. *Clin Gastroenterol Hepatol* 2018;16:1098.e1–105.e1.
14. Weinstein-Nakar I, Focht G, Church P, et al., ImageKids study group. Associations among mucosal and transmural healing and fecal level of calprotectin in children with Crohn's disease. *Clin Gastroenterol Hepatol* 2018;16:1089.e4–97.e4.
15. Cozijnsen MA, Ben Shoham A, Kang B, et al. Development and validation of the mucosal inflammation noninvasive index for pediatric Crohn's disease. *Clin Gastroenterol Hepatol* 2019;18:133.e1–40.e1.
16. 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;46:969–73.
17. Daperno M, Comberlato M, Bossa F, et al., IGIBDEndo Group. Training programs on endoscopic scoring systems for inflammatory bowel disease lead to a significant increase in interobserver agreement among community gastroenterologists. *J Crohns Colitis* 2017;11:556–61.
18. Travis SPL, Schnell D, Krzeski P, et al. Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity (UCEIS). *Gut* 2012;61:535–42.
19. Corte C, Fernandopulle N, Catuneanu AM, et al. Association between the Ulcerative Colitis Endoscopic Index Of Severity (UCEIS) and outcomes in acute severe ulcerative colitis. *J Crohns Colitis* 2015;9:376–81.
20. Ikeya K, Hanai H, Sugimoto K, et al. The Ulcerative Colitis Endoscopic Index of Severity more accurately reflects clinical outcomes and long-term prognosis than the Mayo Endoscopic Score. *J Crohns Colitis* 2016;10:286–95.
21. Xie T, Zhang T, Ding C, et al. Ulcerative Colitis Endoscopic Index of Severity (UCEIS) versus Mayo Endoscopic Score (MES) in guiding the need for colectomy in patients with acute severe colitis. *Gastroenterol Rep (Oxf)* 2018;6:38–44.
22. Schoepfer AM, Beglinger C, Straumann A, et al. Fecal calprotectin correlates more closely with the Simple Endoscopic Score for Crohn's disease (SES-CD) than CRP, blood leukocytes, and the CDAI. *Am J Gastroenterol* 2010;105:162–9.
23. Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2003.
24. Feagan B G, Sandborn W J, D'Haens G, et al. The role of centralized reading of endoscopy in a randomized controlled trial of mesalamine for ulcerative colitis. *Gastroenterology* 2013;145:149.e2–57.e2.
25. Rutgeerts P, Reinisch W, Colombel J-F, 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;83:188.e1–97.e3.
26. Levesque BG, Sandborn WJ, Ruel J, et al. Converging goals of treatment of inflammatory bowel disease from clinical trials and practice. *Gastroenterology* 2015;148:37.e1–51.e1.
27. Nakagawa S, Johnson PCD, Schielzeth H. The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J Royal Soc Interface* 2017;14:20170213.
28. Wu S, Crespi CM, Wong WK. Comparison of methods for estimating the intraclass correlation coefficient for binary responses in cancer prevention cluster randomized trials. *Contemp Clin Trials* 2012;33:869–80.
29. Sikong Y, Lin X, Liu K, et al. Effectiveness of systematic training in the application of narrow-band imaging international colorectal endoscopic (NICE) classification for optical diagnosis of colorectal polyps: experience from a single center in China. *Dig Endosc* 2016;28:583–91.
30. Repici A, Ciscato C, Correale L, et al. Narrow-band Imaging International Colorectal Endoscopic Classification to predict polyp histology: REDEFINE study (with videos). *Gastrointest Endosc* 2016;84:479-486.e3.
31. Sakamoto T, Takamaru H, Sekiguchi M, et al. Reliability of Japan Narrow-Band Imaging Expert Team Classification for the diagnosis of colorectal neoplasms: a pilot study. *Digestion* 2019:1–6.
32. Deganello Saccomani M, Norsa L, Oliva S, et al. The pediatric endoscopy practice in Italy: a nationwide survey on behalf of the Italian Society Of Pediatric Gastroenterology, Hepatology and Nutrition (SIGENP). *Dig Liver Dis* 2019;51:1203–6.

MPG

Manuscript No. JPGN-20-1294

Journal of Pediatric Gastroenterology and Nutrition

Dear Author,

During the preparation of your manuscript for typesetting, some queries have arisen. These are listed below. Please check your typeset proof carefully and mark any corrections in the margin as neatly as possible or compile them as a separate list. This form should then be returned with your marked proof/list of corrections to the Production Editor.

QUERIES: to be answered by AUTHOR

QUERY NO.	QUERY DETAILS	RESPONSE
<AQ1>	Please confirm whether surnames/family names (red) have been identified correctly in the author byline.	
<AQ2>	Affiliations have been set as per style. Please check for accuracy of information.	
<AQ3>	Please check and confirm the corresponding author's name and relevant details for correctness.	
<AQ4>	Please check the abbreviation IC for correctness. Should it be CI?	
<AQ5>	Please check and confirm the corresponding author's name and relevant details for correctness.	
<AQ6>	References have been updated using PubMed. Please check.	
<AQ7>	Please check Table 1 for the accuracy of presentation of the content.	