Revista Española de Enfermedades Digestivas
The Spanish Journal of Gastroenterology

Title:
Interobserver variability of endoscopic scores and their subsections in Crohn's disease: a single-center prospective study

Authors:
Laura Lorenzo González, Daniel Barranco Castro, Victoria Alejandra Jiménez García, Teresa Romero García, Rocío Ferreiro Iglesias, Isabel Vera Mendoza, Blas José Rodríguez Gómez, Manuel Rodríguez-Téllez, Cloé Charpentier, Federico Argüelles-Arias

DOI: 10.17235/reed.2025.10923/2024
Link: PubMed (Epub ahead of print)

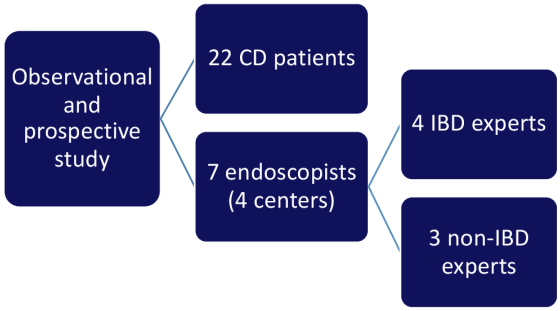
Please cite this article as:
Lorenzo González Laura, Barranco Castro Daniel, Jiménez García Victoria Alejandra , Romero García Teresa, Ferreiro Iglesias Rocío, Vera Mendoza Isabel, Rodríguez Gómez Blas José, Rodríguez-Téllez Manuel, Charpentier Cloé, Argüelles-Arias Federico. Interobserver variability of endoscopic scores and their subsections in Crohn's disease: a single-center prospective study. Rev Esp Enferm Dig 2025. doi: 10.17235/reed.2025.10923/2024.

Revista Española de Enfermedades Digestivas
The Spanish Journal of Gastroenterology

INTEROBSERVER VARIABILITY OF ENDOSCOPIC SCORES AND THEIR SUBSECTIONS IN CROHN'S DISEASE: A SINGLE-CENTER PROSPECTIVE STUDY.

Methods

Objective: Assess interobserver reproducibility of CDEIS and SES-CD in Crohn's disease among IBD experts vs. non-experts.



Key results

Overall ICC: 0.83 (CDEIS), 0.77 (SES-CD) – substantial agreement.

Expert vs Non-expert Agreement:

- Non-experts: ICC 0.91 (CDEIS), 0.88 (SES-CD).
- Experts: ICC 0.79 for both.

Low agreement for deep ulcers in specific areas and ulcer size.

Fecal Calprotectin: Significant correlation with SES-CD (r=0.58, p=0.01).
No correlation with **Harvey-Bradshaw index/CRP.**

Conclusions

•CDEIS and SES-CD are reproducible even without specialized training.

•Both scores have limitations detecting mild inflammation, particularly ileal activity and ulcer severity.

•Highlighting ileal activity and simplifying ulcer and stricture assessment may enhance clinical applicability.

Lorenzo Gonzalez, et al.

Revista Española de Enfermedades Digestivas (REED)
The Spanish Journal of Gastroenterology

Revista Española de Enfermedades Digestivas
The Spanish Journal of Gastroenterology

Revista Española de Enfermedades Digestivas
The Spanish Journal of Gastroenterology

Interobserver variability of endoscopic scores and their subsections in Crohn's Disease: a single-center prospective study

Laura Lorenzo Gonzalez[1], Daniel Barranco Castro[1], Victoria Alejandra Jimenez-Garcia[1], Teresa Romero García[2], Rocío Ferreiro Iglesias[3], Isabel Vera Mendoza[4], Blas Rodríguez Gómez[1], Manuel Rodrigez-Tellez[1], Cloé Charpentier[5], Federico Argüelles Arias[1]

1. Department of Gastroenterology, Virgen Macarena University Hospital, Seville, Spain.
2. Department of Gastroenterology, San Juan de Dios del Aljarafe Hospital, Seville, Spain.
3. Department of Gastroenterology, University Hospital Santiago de Compostela, Galicia, Spain.
4. Department of Gastroenterology, Puerta de Hierro University Hospital, Majadahonda, Spain.
5. Department of Gastroenterology, Rouen University Hospital, Rouen, France.

Address for correspondence:

Laura Lorenzo Gonzalez.

Flat 1, 182 Chiswick High Road, W4 1PP, London. UK.

Email: lauralorenzog93@gmail.com

Conference presentation:

Poster published in ESGE Days, Berlin, 2024. A brief summary of the work was presented.

Abbreviations:

IBD: Inflammatory bowel disease**, CD:** Crohn's Disease**, CDEIS:** Crohn's Disease endoscopic index of severity, **SES-CD:** simple endoscopic score for Crohn's Disease, **ER** endoscopic

Revista Española de Enfermedades Digestivas
The Spanish Journal of Gastroenterology

remission, **CRP**: C-reactive protein, **FCP**: fecal calprotectin, **SD**: standard deviation**, IQR**: interquartile range, **ICC**: intra-class correlation coefficient.

Abstract

Background and Aims: Accurate endoscopic assessment is crucial for treatment decisions in Crohn's disease (CD), but endoscopic scores are complex and not easily applicable in routine practice. This study aimed to assess interobserver reproducibility of CDEIS, SES-CD and their subsections, in inflammatory bowel disease (IBD) experts and non-experts, to improve endoscopic assessment.

Methods: Observational, prospective study including 22 CD patients who underwent routine colonoscopy at an IBD unit, after excluding patients with inadequate bowel preparation (Boston Bowel Preparation score <6). 7 endoscopists from 4 centers, 4 specialized in IBD, independently scored the videos using CDEIS and SES-CD. Inter-observer variability was assessed, comparing IBD experts and non-experts, and correlating endoscopic scores with clinical activity and biomarkers.

Results: Overall intraclass correlation coefficient (ICC) was 0.83 for CDEIS and 0.77 for SES-CD, indicating substantial agreement. The lowest correlations were deep ulcers in ileum, descending colon, and rectum (CDEIS), and ulcer size in ileum and stricture detection in descending colon (SES-CD). Non-IBD experts showed higher interobserver agreement (ICC: 0.91 CDEIS, 0.88 SES-CD) compared to IBD experts (ICC: 0.79 for both). No correlation was found between endoscopic scores and Harvey-Bradshaw index or CRP. SES-CD showed significant correlation with fecal calprotectin (0.58, p=0.01) and CDEIS trended towards significance (r=0.44, p=0.064). 22% patients were inconsistently classified regarding endoscopic remission.

Conclusions: Although CDEIS and SES-CD are highly reproducible without specialized training, they have limitations detecting mild inflammatory activity. An ideal score should emphasize ileal activity and simplify assessment of ulceration severity and stenosis to improve clinical utility.

Keywords: Crohn’s Disease. Endoscopic scores. Interobserver variability.

Revista Española de Enfermedades Digestivas
The Spanish Journal of Gastroenterology

Introduction

Over the past few decades, the treatment of Crohn's disease (CD) has significantly advanced, with mucosal healing emerging as a central therapeutic goal linked to better long-term outcomes. Endoscopic remission (ER), defined as the absence of ulceration, has become a key target [1,2,3]. Ileocolonoscopy remains the gold standard for assessing CD inflammation, with the CD Endoscopic Index of Severity (CDEIS) and the Simple Endoscopic Score for CD (SES-CD) being widely used scoring systems[4]. Although SES-CD is simpler and faster than CDEIS, both scores have limitations, including underestimating strictures and overestimating non-specific lesions[4,5]. These indices were developed when treatment options were more limited, focusing on lesion severity, but the current therapeutic focus on mucosal healing requires more precise assessment, particularly for mild lesions[6]. Reproducibility of these scores is crucial, as variability can affect treatment outcomes and trial eligibility. While prior studies have shown good reproducibility for both scores, limited data exists on the reproducibility of their subsections or the impact of IBD expertise [7,8,9].

Methods

We conducted a prospective, observational, single-center study at Virgen Macarena University Hospital, Seville, including adult outpatients with histological diagnosis of CD, who underwent a colonoscopy in their routine management between November 2019 and November 2021. Patients with poor video quality or inadequate bowel preparation (Boston scale[10] <6) were excluded. Colonoscopies were recorded using the AVerMedia EZRecorder 330 device and anonymized for review by 7 endoscopists, who were blinded to clinical and treatment details. The panel included 4 endoscopists from our center (1 IBD expert), and 3 IBD specialists from other national centers. Each endoscopist independently scored the videos using CDEIS and SES-CD. The endoscopists had an experience of at least 15 years or more than 10000 colonoscopies and IBD experts were defined as endoscopists with specialized IBD training, weekly IBD consultation, or affiliation with a national IBD unit. ER

was defined as a CDEIS <3 or SES-CD =2.[11]

Clinical data were obtained from computerized medical records, with the Harvey-Bradshaw Index (HBI) calculated using data collected within 3 months before or after colonoscopy. Biomarkers, including C-reactive protein (CRP) and fecal calprotectin (FCP), were analyzed when available within this time frame. The indication for colonoscopy was not identified in our work.

Our main objective was to analyze the interobserver variability of CDEIS and SES-CD and their subsections, in IBD experts and non-IBD experts. Our secondary objective was to analyze the correlation between each endoscopic scale and clinical activity score or inflammatory biomarkers.

Statistical analysis was conducted using IBM SPSS Statistics 24. Categorical variables were presented as percentages, and continuous variables as means with standard deviations or medians with interquartile ranges, depending on normality. Concordance between endoscopists' scores was assessed using kappa indices for dichotomous and categorical variables, and intra-class correlation coefficients (ICC) for continuous variables. Concordance strength was classified as fair (≤0.40), moderate (0.41-0.60), substantial (0.61-0.80), or almost perfect (>0.80). Linear correlations were analyzed using Pearson or Spearman coefficients based on distribution. A p-value ≤0.05 was considered statistically significant, and a subgroup analysis compared IBD experts to non-experts.

This study was approved by the research ethics committee of the university hospital involved. The patients were included after providing informed written consent.

Results

Characteristics of patients

A total of 25 patients were enrolled, with 3 excluded due to poor video quality or poor bowel preparation, leaving 22 patients for analysis. The cohort had majority of men, a mean

age of 40 years, and predominant presentation of ileal and stricturing disease. Five patients had prior IBD surgery. Colonoscopy reached the terminal ileum in 17 patients, ileocecal valve in 3, and right colon in 2, with all cases of incomplete intubation due to strictures.

Bowel preparation scores on the Boston scale were 9 (16 patients), 8 (1 patient), 7 (1 patient), and 6 (4 patients). CRP and fecal calprotectin data were available for 17 and 18 patients, respectively.

Table 1. Characteristics of patients.

N=22		
Age (years)	Mean (SD)	40.6 (+/- 13.3)
Male gender	N (%)	13 (59.1%)
Duration of CD (years)	Median (IQR)	8.5 (3,0 – 17.3)
CD localization: - Ileal L1 - Colonic L2 - ileocolonic L3	N (%)	 12 (54.5%) 2 (9.1%) 8 (36.4%)
CD behavior: - inflammatory B1 - stricturing B2 - penetrating B3	N (%)	 10 (45.5%) 11 (50.0%) 1 (4.5%)
Perianal localization	N (%)	2 (9.1%)
Prior surgery	N (%)	5 (22.7%)
Ongoing treatment: - 5-ASA - steroids - thiopurines - biologics	N (%)	 6 (27.3%) 2 (9.1%) 6 (27.3%) 8 (36.4%)
Harvey-Bradshaw index	Median	5 (3 – 8)

Revista Española
de Enfermedades Digestivas
The Spanish Journal
of Gastroenterology

	(IQR)	

Revista Española
de Enfermedades Digestivas
The Spanish Journal
of Gastroenterology

	Median (IQR)	
CRP (mg/L)	Median (IQR)	8.15 (2.9 – 18.7)
FCP (μg/g)	Median (IQR)	653.15 (177 – 1242)

CD = Crohn's disease. SD = standard deviation. IQR = interquartile range. 5-ASA = 5-aminosalicylic acid. FCP= Fecal calprotectine.

Results of CDEIS and SES-CD

Overall median of CDEIS was 7.27 (3.65 – 11) and SES-CD was 7 (4 – 11). Considering the median total score for each patient separately, 4 (18.2%) and 3 (13.6%) of them completed the definition of ER according to evaluation with CDEIS and SES-CD, respectively.

Interobserver agreement was higher among internal evaluators for CDEIS (ICC 0.91; 95% CI: 0.84–0.96) compared to external evaluators (ICC 0.73; 95% CI: 0.54–0.87), while SES-CD showed comparable agreement between internal (ICC 0.74; 95% CI: 0.58–0.87) and external evaluators (ICC 0.78; 95% CI: 0.62–0.90).

In our study, 5 of 22 patients (22.7%) had discordant CDEIS scores, with some evaluators classifying them as in remission (CDEIS <3) and others as having active disease. Similarly, SES-CD scores differed in 4 of 22 patients (18.2%), leading to disagreement on remission status. Although neither index has validated cut-off values, remission is commonly defined in the literature as CDEIS <3 and SES-CD =2.

Interobserver correlation for CDEIS

The global ICC for the 7 endoscopists was 0.83 (0.733 – 0.915). Table 2 illustrates the analysis of each subsection. The lowest correlation coefficients (≤0,40) corresponded to detection of deep ulcers in ileum, descending colon and rectum, as well as quantification of ulcerated or affected surface in ileum. On the opposite, the only correlation coefficient reaching >0.80

corresponded to the detection of superficial ulcers in the rectum. When grouping together the different segments for each item, superficial ulcers and quantification of affected surface showed the higher agreement.

In the sub-group analysis, non-IBD experts showed higher ICC values for CDEIS (0.91) and SES-CD (0.88) compared to IBD experts (0.79 for both indices), indicating better overall agreement. Non-IBD experts had stronger agreement in detecting deep ulcers in rectum, superficial ulcers in colon and quantifying affected surface in colon, while IBD experts had higher ICC for non-ulcerated stenosis.

Table 2. Correlation analysis of subsections of the CDEIS.

	Global		**IBD experts**		**Non- IBD experts**	
	Mean of Kappa ponderated	**ICC**	**Mean of Kappa ponderated**	**ICC**	**Mean of Kappa ponderated**	**ICC**
Deep ulcers						
- overall	0.46		0.42		0.49	
- ileum	0.31		0.39		0.23	
- ascending colon	0.46		0.43		0.38	
- transverse colon	0.71		0.63		0.80	
- descending colon	0.27		0.04		0.52	
- rectum	0.20		0		0.33	
Superficial ulcers						
- overall	0.62		0.54		0.75	
- ileum	0.46		0.47		0.47	

- ascending colon	0.49		0.36		0.71	
- transverse colon	0.67		0.51		0.83	
- descending colon	0.72		0.59		0.91	
- rectum	0.84		0.83		0.85	
Ulcerated surface						
- overall		0.73				
- ileum		0.32		0.47		0.58
- ascending colon		0.74		0.66		0.79
- transverse colon		0.62		0.49		0.76
- descending colon		0.57		0.372		0.79
- rectum		0.72		0.60		0.87
Affected surface						
- overall		0.83				
- ileum		0.38		0.44		0.54
- ascending colon		0.72		0.59		0.89
- transverse colon		0.70		0.56		0.90
- descending colon		0.64		0.46		0.91
- rectum		0.80		0.90		0.77
Ulcerated stenosis	**0.59**		**0.60**		**0.62**	

Non-ulcerated stenosis	**0.60**		**0.66**		**0.45**	
Total CDEIS CI 95%	**0.84** (0.73 – 0.91)		**0.79** (0.65-0.89)		**0.91** (0.82-0.96)	

ICC = intra-class correlation coefficient. CI = confidence interval. The best (>0.80) and worst (≤0.40) values are highlighted in green and orange.

Interobserver correlation for SES-CD

The global ICC for the 7 endoscopists was 0.77 (0.64–0.88). Results in table 3. The lowest correlations were observed for ulcer size and quantification of ulcerated or affected surface in the ileum, as well as stricture detection in the descending colon, with none exceeding an ICC of 0.80. When grouped by segment, the highest ICC was for stricture detection (0.68).

In the sub-group analysis, both expert groups had similar ICC, with non-IBD experts showing higher agreement on ulcer size in the transverse and descending colon, and IBD experts on stenosis in ascending colon.

Table 3. Correlation analysis of subsections of the SES-CD.

	Global		**IBD experts**		**Non-IBD experts**	
	Mean of Kappa ponderated	**ICC**	**Mean of Kappa ponderated**	**ICC**	**Mean of Kappa ponderated**	**ICC**
Size of ulcers						

- overall	0.58		0.56		0.63	
- ileum	0.40		0.51		0.42	
- ascending colon	0.46		0.43		0.41	
- transverse colon	0.69		0.58		0.81	
- descending colon	0.61		0.52		0.78	
- rectum	0.68		0.68		0.75	
Ulcerated surface						
- overall	0.57		0.59		0.62	
- ileum	0.31		0.47		0.36	
- ascending colon	0.57		0.58		0.46	
- transverse colon	0.66		0.56		0.76	
- descending colon	0.61		0.58		0.74	
- rectum	0.73		0.74		0.72	
Affected surface						
- overall	0.59		0.60		0.66	
- ileum	0.36		0.44		0.46	
- ascending colon	0.50		0.43		0.56	
- transverse colon	0.71		0.77		0.73	
- descending colon	0.67		0.59		0.80	
- rectum	0.76		0.80		0.66	

Revista Española de Enfermedades Digestivas
The Spanish Journal of Gastroenterology

Stenosis						
- overall	0.68		0.70		0.65	
- ileum	0.65		0.67		0.65	
- ascending colon	0.59		0.68		0.40	
- transverse colon	0.47		0.50		0.33	
- descending colon	0		0		-	
- rectum	-		-		-	
Total SES-CD CI 95%		**0.77** (0.64–0.88).		**0.79** (0.65-0.89)		**0.88** (0.78-0.94)

ICC = intra-class correlation coefficient. - = not calculable because there is no variability due to a single matching response. The best (>0.80) and worst (≤0.40) values are highlighted in green and orange.

Correlation between endoscopic scores, clinical score and inflammatory biomarkers

The Spearman coefficient was low for both CDEIS and SES-CD with HBI and with CRP, illustrating non-significant correlation (Table 4). On the other hand, the Spearman coefficient was 0,445 for CDEIS and FCP (p=0,064), and 0,582 for SES-CD and FCP, reaching statistical significance (p=0,011).

Table 4. Correlation analysis of endoscopic scores with clinical score and inflammatory biomarkers.

	HBI (N = 22)		**CRP** (N = 17)		**FCP** (N = 18)	
	Spearman coefficient	p	Spearman coefficient	p	Spearman coefficient	p
Median CDEIS	0.09	0.68	0.23	0.35	0.44	0.06
Median SES-CD	0.11	0.64	0.02	0.45	0.58	**0.01**

CDEIS = Crohn's disease endoscopic index of severity. SES-CD = simple endoscopic score for Crohn's disease. HBI = Harvey-Bradshaw index. CRP = C-reactive protein. FCP = fecal calprotectin.

Discussion

CDEIS and SES-CD demonstrated good concordance in our work, though their subsections showed disparate correlation scores. As expected,[4,12–14] endoscopic scores for CD had poor correlation with HBI, with only FCP showing a significant correlation with SES-CD.

A recent meta-analysis [15] reported overall agreement rates of 0.80 for CDEIS and 0.78 for SES-CD, closely aligning with our findings and reflecting previous literature with similar rates of interobserver agreement.[7,16,17] Several studies have examined score subsections, with Daperno et al. [17] highlighting variability in CDEIS for superficial ulcers[4], and subsequent research showing poorer reproducibility for SES-CD in assessing affected surface and all items in the ascending colon.[16]Some authors also concluded that interpreting ulcer depth, estimating surface area, defining anatomical location of lesions, differentiating between anal and rectal lesions, or grading severity of stenosis, were sources of disagreement.[7,18] Our findings are consistent with these observations. We found the greatest variability in scoring the ulcerated surface, followed by ulcer size and affected surface.

Higher variability in ileum assessments is likely due to difficulties in positioning and peristaltic contractions.[18] With 91% of our patient showing ileal involvement, greater reproducibility is expected when endoscopists assess the absence of lesions rather than their severity .This is consistent with Hart et al,[19] who observed higher agreement on identifying healed bowel or severe disease compared to mild-to-moderate disease. The anatomical location of the lesions can also be challenging when using endoscopic scores,[7,18] especially in post-inflammatory remodeling, after surgery or in lesions involving contiguous segments. In our work we observed greater correlation by grouping segments for affected or ulcerated surface. In practice, the value of separating colonic segments appears limited, as it proves laborious and has no therapeutic impact.

Similarly, assessing ulcer depth remained a challenge despite the introduction of standardized terminologies by the ECCO committee in 2013, aimed at providing clearer definitions of shallow versus deep ulcers.[20] Despite these efforts, distinguishing between the two remains difficult, with conflicting results in the literature. [4,7,16,18]

In our study, we found no evidence that IBD expertise could lead to better reproducibility of CD endoscopic scores, except perhaps slightly better results for the detection of stenoses. Some authors previously reported that gastroenterologists and trainees performed similar agreement in SES-CD, without specific formation.[19]

In terms of clinical utility, CDEIS and SES-CD were developed in a very different context, when no effective treatment was available for IBD. The objective at that time was to elaborate a standardized, reproducible index to quantify the lesion severity, reflecting the global endoscopists' assessment, along the entire range of severity. [21, 6] Our current target of mucosal healing calls for better discrimination with respect to mild lesions.[22] In our work, despite good reproducibility, over 20% of patients had discrepancies in ER assessment (SES-CD <2 and CDEIS ≤3[6,24]) among the 7 endoscopists.

We propose that an ideal scoring should emphasize ileal involvement, which is predictive of poorer outcomes and is currently underestimated by the existing 5-segment scale.

Simplifying the score by eliminating the need to separate colonic segments and focusing on a single marker of ulceration severity (size, depth, or surface area) would enhance its utility. The role of stenosis should be reassessed, as detection and graduation of stenosis are sources of interobserver variability.

In the past few years, the aims of CD endoscopic scores have shifted towards prognostic and predictive roles, as they are supposed to guide our therapeutic management. SES-CD have been the subject of recent publications on their predictive capacity during CD treatment.[23] It is widely accepted that more severe endoscopic activity is predictive of poorer long-term CD evolution,[24] but it has been shown that rather than the overall degree of baseline inflammation, ileal and rectal involvement or deep ulcers are associated with poorer clinical outcome or lower rate of ER.[25–27]

Recent publications have attempted to imbue the scores with clinical utility and simplify them as Narula et al. with the Modified Multiplier SES-CD (MM-SES-CD), [28] which assigns individual weights to each SES-CD parameter based on its impact on achieving endoscopic remission and Adler et al. and the Simplified Endoscopic Mucosal Assessment for CD (SEMA-CD), [29] which is much easier to use, following the example of the Mayo or Rutgeerts endoscopic scores.

Our present study suffers few limitations. We included a limited number of patients, with a strong predominance of ileal location and stricturing disease, so that our results might lack statistical power and not be extrapolable to any CD cohort. We didn't perform longitudinal follow-up, so we were unable to assess sensitivity to change and predictive performance of CDEIS and SES-CD. However, this is a comprehensive work on CD endoscopic scores and their subsections, conducted on real-life patients outside therapeutic trials, with 7 endoscopists from 4 different national centers, which adds to a very sparse literature.

In conclusion, our study confirms that both CDEIS and SES-CD are highly reproducible without the need for specialised expertise. However, they may have difficulty in discriminating mild inflammatory activity. We suggest that an ideal scoring system should

prioritise ileal involvement, which predicts worse outcomes but is often underestimated. Simplifying the score by eliminating the need to separate colonic segments and focusing on a single defined marker of ulceration severity (size or depth) and reassessing stenosis, which causes interobserver variability. These changes can improve clinical utility and reduce interobserver variability.

Conflicts of interest

All authors participating confirm that they have no conflicts of interest to declare.

Funding

No specific funding has been received to elaborate this study.

Data Availability Statement

The data underlying this article cannot be shared publicly as it contains clinical and analytical data of patients, and the privacy of individuals that participated in the study must be maintained.

Bibliography

1. Ungaro RC, Yzet C, Bossuyt P, et al. Deep Remission at 1 Year Prevents Progression of Early Crohn's Disease. *Gastroenterology*. 2020;159(1):139-147. doi:10.1053/j.gastro.2020.03.039
2. Frøslie KF, Jahnsen J, Moum BA, et al. IBSEN Group. Mucosal healing in inflammatory bowel disease: results from a Norwegian population-based cohort. *Gastroenterology*. 2007;133(2):412-422. doi:10.1053/j.gastro.2007.05.051
3. Baert F, Moortgat L, Van Assche G, et al. Mucosal healing predicts sustained clinical remission in patients with early-stage Crohn's disease. *Gastroenterology*. 2010;138(2):463-468; quiz e10-11. doi:10.1053/j.gastro.2009.09.056

4. 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):505-512. doi:10.1016/s0016-5107(04)01878-4
5. Sipponen T, Nuutinen H, Turunen U, et al. Endoscopic evaluation of Crohn's disease activity: comparison of the CDEIS and the SES-CD. *Inflamm Bowel Dis*. 2010;16(12):2131-2136. doi:10.1002/ibd.21300
6. Vuitton L, Marteau P, Sandborn WJ, et al. IOIBD technical review on endoscopic indices for Crohn's disease clinical trials. *Gut*. 2016;65(9):1447-1455. doi:10.1136/gutjnl-2015-309903
7. 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):1119-1125. doi:10.1136/gutjnl-2014-308973
8. 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):1006-1014. doi:10.1093/ecco-jcc/jjw120
9. 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):969-973. doi:10.1016/j.dld.2014.07.010
10. Lai EJ, Calderwood AH, Doros G, et al. The Boston bowel preparation scale: a valid and reliable instrument for colonoscopy-oriented research. *Gastrointest Endosc*. 2009;69(3 Pt 2):620-625. doi:10.1016/j.gie.2008.05.057
11. Vuitton L, Marteau P, Sandborn WJ, et al. IOIBD technical review on endoscopic indices for Crohn's disease clinical trials. *Gut*. 2016;65(9):1447-1455. doi:10.1136/gutjnl-2015-309903
12. Sostegni R, Daperno M, Scaglione N, et al. Review article: Crohn's disease: monitoring disease activity. *Aliment Pharmacol Ther*. 2003;17 Suppl 2:11-17. doi:10.1046/j.1365-2036.17.s2.17.x
13. Sipponen T, Nuutinen H, Turunen U, et al. Endoscopic evaluation of Crohn's disease activity: comparison of the CDEIS and the SES-CD. *Inflamm Bowel Dis*. 2010;16(12):2131-2136. doi:10.1002/ibd.21300

14. Jones J, Loftus EV, Panaccione R, et al. Relationships between disease activity and serum and fecal biomarkers in patients with Crohn's disease. *Clin Gastroenterol Hepatol*. 2008;6(11):1218-1224. doi:10.1016/j.cgh.2008.06.010
15. Hashash JG, Yu Ci Ng F, Farraye FA, et al. Inter- and Intraobserver Variability on Endoscopic Scoring Systems in Crohn's Disease and Ulcerative Colitis: A Systematic Review and Meta-Analysis. *Inflamm Bowel Dis*. 2024;30(11):2217-2226. doi:10.1093/ibd/izae051
16. 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):188-197.e1-3. doi:10.1016/j.gie.2015.06.018
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*. 2017;11(5):556-561. doi:10.1093/ecco-jcc/jjw181
18. 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):1006-1014. doi:10.1093/ecco-jcc/jjw120
19. Hart L, Chavannes M, Lakatos PL, et al. Do You See What I See? An Assessment of Endoscopic Lesions Recognition and Description by Gastroenterology Trainees and Staff Physicians. *J Can Assoc Gastroenterol*. 2020;3(5):216-221. doi:10.1093/jcag/gwz022
20. Annese V, Daperno M, Rutter MD, et al. European evidence based consensus for endoscopy in inflammatory bowel disease. *Journal of Crohn's and Colitis*. 2013;7(12):982-1018. doi:10.1016/j.crohns.2013.09.016
21. 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(7):983-989.
22. Neurath MF, Vieth M. Different levels of healing in inflammatory bowel diseases: mucosal, histological, transmural, barrier and complete healing. *Gut*. 2023;72(11):2164-2183. doi:10.1136/gutjnl-2023-329964

23. Khanna R, Zou G, Stitt L, et al. Responsiveness of Endoscopic Indices of Disease Activity for Crohn's Disease. *Am J Gastroenterol*. 2017;112(10):1584-1592. doi:10.1038/ajg.2016.580
24. Khanna R, Bouguen G, Feagan BG, et al. A systematic review of measurement of endoscopic disease activity and mucosal healing in Crohn's disease: recommendations for clinical trial design. *Inflamm Bowel Dis*. 2014;20(10):1850-1861. doi:10.1097/MIB.0000000000000131
25. Allez M, Lemann M, Bonnet J, et al. Long term outcome of patients with active Crohn's disease exhibiting extensive and deep ulcerations at colonoscopy. *Am J Gastroenterol*. 2002;97(4):947-953. doi:10.1111/j.1572-0241.2002.05614.x
26. Narula N, Wong ECL, Aruljothy A, et al. Ileal and Rectal Ulcer Size Affects the Ability to Achieve Endoscopic Remission: A Post hoc Analysis of the SONIC Trial. *Am J Gastroenterol*. 2020;115(8):1236-1245. doi:10.14309/ajg.0000000000000617
27. Rivière P, D'Haens G, Peyrin-Biroulet L, et al. Location but Not Severity of Endoscopic Lesions Influences Endoscopic Remission Rates in Crohn's Disease: A Post Hoc Analysis of TAILORIX. *Am J Gastroenterol*. 2021;116(1):134-141. doi:10.14309/ajg.0000000000000834
28. Narula N, Wong ECL, Colombel JF, et al. Predicting endoscopic remission in Crohn's disease by the modified multiplier SES-CD (MM-SES-CD). *Gut*. 2022;71(6):1078-1087. doi:10.1136/gutjnl-2020-323799
29. Adler J, Colletti RB, Noonan L, et al. Validating the Simplified Endoscopic Mucosal Assessment for Crohn's Disease: A Novel Method for Assessing Disease Activity. *Inflamm Bowel Dis*. 2023;29(7):1089-1097. doi:10.1093/ibd/izac183

Authors' contribution

L.G.L.: study design, patient recruitment, data collection, data analysis, writing.

B. C. D.: patient recruitment, data collection.

J.G.V.A.: recordings evaluation.

R.G.T.: recordings evaluation.

F.I.R.: recordings evaluation.

Revista Española de Enfermedades Digestivas
The Spanish Journal of Gastroenterology

V.M.I.: recordings evaluation.

R.G.B.: recordings evaluation.

R.T.M.: recordings evaluation.

C.C.: writing.

A.A.F.: study design, recordings evaluation, writing, coordination.