



Original Article



Interrater Reliability of the Nancy Histologic Index in Assessing Histologic Remission in Treated Ulcerative Colitis Biopsies: A Multi-institutional Experience Among Gastrointestinal Pathologists in the United States

Krithika D. Shenoy¹, Jiannan Li¹, Daniela Allende², Samuel J. Ballentine¹, Kathleen Byrnes¹, Parakkal Deepak³, Alicia G. Dessain⁴, Ashwini K. Esnakula⁵, Raul S. Gonzalez⁶, Xianyong Gui⁷, Hwajeong Lee⁸, Jingmei Lin⁷, Shivani Mattay⁹, Namrata Setia¹⁰, Hanlin L. Wang¹¹, Zhaohai Yang¹², Xuchen Zhang¹³, Xiuli Liu^{1*} and on behalf of the SPARC-IBD Investigators

¹Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO, USA; ²Department of Pathology, Robert J Tomsich Pathology and Laboratory Medicine Institute, Cleveland Clinic Foundation, Cleveland, OH, USA; ³Division of Gastroenterology, Washington University School of Medicine, St. Louis, MO, USA; ⁴Department of Pathology and Anatomical Sciences, University of Missouri School of Medicine, Columbia, MO, USA; ⁵Department of Pathology, The Ohio State University Wexner Medical Center/James Cancer Hospital, Columbus, OH, USA; ⁶Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA, USA; ⁷Department of Pathology, Wake Forest University School of Medicine, Winston-Salem, NC, USA; ⁸Department of Pathology, Albany Medical College, Albany, NY, USA; ⁹John T. Milliken Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA; ¹⁰Department of Pathology, The University of Chicago, Chicago, IL, USA; ¹¹Department of Pathology and Lab Medicine, David Geffen School of Medicine at the University of California Los Angeles, Los Angeles, CA, USA; ¹²Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA; ¹³Department of Pathology, Yale School of Medicine, New Haven, CT, USA

Received: April 20, 2025 | Revised: May 29, 2025 | Accepted: June 05, 2025 | Published online: June 26, 2025

Abstract

Background and objectives: Histologic remission is recommended as an adjunctive treatment target in ulcerative colitis, and scoring systems have been proposed to enhance reproducibility. The Nancy Histologic Index (NHI) is increasingly used in clinical trials; however, its performance in real-world settings is not fully established. This study aimed to assess the interrater reliability (IRR) of the NHI among gastrointestinal pathologists in the United States. **Methods:** Thirty-seven whole-slide images of colorectal biopsies from 34 treated ulcerative colitis patients enrolled in a multicenter adult cohort were independently reviewed by 12 gastrointestinal pathologists. Each biopsy was reviewed twice, five months apart, and graded using the NHI. Prior to the second review, pathologists completed an online tutorial on the NHI. **Results:** The NHI showed substantial IRR in both reviews [intraclass correlation coefficient (ICC) = 0.79; 95% confidence interval (CI), 0.70–0.87 at Review 1; ICC = 0.78; 95% CI, 0.69–0.86 at Review 2]. However, considerable variability was observed in individual grade assignments, with the lowest IRR for Grade

2 (ICC = 0.24; 95% CI, 0.15–0.37; $P < 0.001$, and ICC = 0.23; 95% CI, 0.14–0.36; $P < 0.001$ for Reviews 1 and 2, respectively), followed by Grade 4 (ICC = 0.41; 95% CI, 0.29–0.55; $P < 0.001$, and ICC = 0.47; 95% CI, 0.35–0.61; $P < 0.001$). Grade 1 showed the highest IRR (ICC = 0.79; 95% CI, 0.70–0.87; $P < 0.001$, and ICC = 0.78; 95% CI, 0.69–0.86; $P < 0.001$). When Grades 2, 3, and 4 (i.e., active disease) were grouped together, the IRR remained substantial across both reviews (ICC = 0.76; 95% CI, 0.66–0.85; $P < 0.001$). **Conclusions:** While the substantial IRR for active disease (Grades ≥ 2) in this study underscores the clinical utility of the NHI, refinement of criteria for Grades 2, 3, and 4 will be crucial in reducing variability among observers and enabling more accurate monitoring of treatment endpoints.

Citation of this article: Shenoy KD, Li J, Allende D, Ballentine SJ, Byrnes K, Deepak P, et al. Interrater Reliability of the Nancy Histologic Index in Assessing Histologic Remission in Treated Ulcerative Colitis Biopsies: A Multi-institutional Experience Among Gastrointestinal Pathologists in the United States. J Clin Transl Pathol 2025. doi: 10.14218/JCTP.2025.00022.

Keywords: Colorectal biopsy; Erosion; Histologic remission; Inflammatory bowel disease; Interrater reliability; Nancy histologic index; Ulcer; Ulcerative colitis.

*Correspondence to: Xiuli Liu, Department of Pathology and Immunology, Washington University in St. Louis School of Medicine, 660 Euclid Ave, St. Louis, MO 63122, USA. ORCID: <https://orcid.org/0000-0001-5791-2017>. Tel: +1-314-273-4843, Fax: +1-314-747-4392, E-mail: l.xiuli@wustl.edu

Introduction

Emerging data suggest that patients with ulcerative colitis (UC) who exhibit persistent histologic activity are at elevated

risk for both short-term and long-term complications, including higher rates of relapse, hospitalization, surgery, and neoplasia, even when apparent endoscopic and clinical remission is achieved.¹⁻³ In light of these findings, the Selecting Therapeutic Targets in Inflammatory Bowel Disease II initiative recommends incorporating histologic remission as an adjunct to endoscopic remission to achieve deeper disease control, consistent with its treat-to-target approach.⁴ While histologic remission is increasingly being adopted as a secondary treatment target in randomized clinical trials and observational studies, the lack of a universally accepted definition complicates its application in routine practice. This poses significant challenges for clinicians seeking to optimize treatment strategies and improve patient outcomes.⁵

Several histologic scoring systems have been proposed over the years to address the limitations of descriptive pathology reports, which often lack standardization and comparability. Among these systems, the Robarts Histopathology Index, Nancy Histologic Index (NHI), and Geboes Score have undergone the most extensive validation and are increasingly used in clinical trials.⁶ However, there remains a paucity of data supporting their use in everyday clinical practice. In a global survey of gastroenterologists and pathologists, 77% of respondents reported that a standardized histologic score was not included in their pathology reports, in contrast to more than 90% who reported using a standardized endoscopic scoring system, such as the Mayo Endoscopic Score (MES), in clinical practice.⁷

The NHI involves a stepwise evaluation of three components: ulceration, acute inflammation, and chronic inflammation. These parameters are used to assign a five-tier grade: Grade 0 (no or mild chronic inflammation), Grade 1 (moderate to severe chronic inflammation), Grade 2 (rare or few neutrophils in the lamina propria or epithelium that are difficult to detect), Grade 3 (multiple clusters of neutrophils in the lamina propria and/or epithelium that are apparent), and Grade 4 (presence of ulceration).^{8,9} It is one of two indices currently recommended by the European Crohn's and Colitis Organisation for use in clinical practice, clinical trials, and observational studies due to its validity and simplicity.¹⁰ However, its reproducibility in real-world settings has not been thoroughly evaluated. The aim of this study was to assess the performance of the NHI among gastrointestinal pathologists at tertiary care academic centers in the United States, using colorectal biopsies from a prospective adult cohort of treated UC patients.

Materials and methods

Case selection and histologic assessment

Thirty-seven hematoxylin and eosin (H&E)-stained whole-slide images of colorectal biopsies from 34 patients with UC who had received treatment and were enrolled in a multicenter longitudinal cohort of the Study of a Prospective Adult Research Cohort with Inflammatory Bowel Disease (SPARC-IBD) were included. Cases were selected from this larger cohort to ensure a heterogeneous distribution from different anatomic sites (right colon, left colon, sigmoid colon, and rectum) and macroscopic appearances during colonoscopy. Endoscopic evaluation was based on MES of 0–3.¹¹ Biopsy samples were collected in formalin using the SPARC-IBD protocol. Biopsies were formalin-fixed, paraffin-embedded, and stained with H&E at a central biobank (Sampled: <https://sampled.com/>). All slides were scanned using an Olympus V120 Virtual Slide Microscope at 40× magnification. The .vsi output images were then converted to .jpg files for easier ac-

cessibility while preserving image quality.

Twelve pathologists with subspecialty training in gastrointestinal pathology, all practicing at tertiary care academic centers in the United States, reviewed each biopsy twice, with a five-month washout interval between reviews to minimize recall bias. Each pathologist evaluated the biopsies using the NHI, as previously described.^{8,9} Pathologists were instructed to treat biopsies with erosion similarly to those with ulceration—that is, to assign them an NHI Grade 4 at both reviews. They also assessed two additional parameters: crypt architectural distortion and Paneth cell metaplasia in biopsies from the left colon and rectum. Pathologists were informed of the anatomic location of each biopsy to account for regional differences in the chronic inflammatory gradient in the normal colon, though the decision to use this information was left to the discretion of each reader. All pathologists were otherwise blinded to clinical and endoscopic data. They had access to the original paper describing and validating the NHI for self-reading but did not receive any pre-study group training before the first review.⁸

The same set of 37 H&E-stained colorectal biopsies was rearranged in a different order and reviewed a second time after five months. Prior to this second round, the pathologists received additional training via an interactive online web tutorial (Supplementary Material 1), curated by authors KDS, JNL, and XL, to address any knowledge gaps. The study pathologists recorded the same histologic parameters as in the first review and remained blinded to the clinical and endoscopic data, except for biopsy site.

Statistical analysis

Statistical analysis was performed using R software, version 4.3.1 for MacOS (R Foundation for Statistical Computing, Vienna, Austria). Inter- and intra-rater agreements and reliability were computed using the irr package. Inter-rater reliability (IRR) for all raters was calculated using the intraclass correlation coefficient (ICC) with a two-way random effects model for absolute agreement. Intra-rater agreement was calculated using unweighted Cohen's kappa. ICC and kappa values were interpreted using the categories proposed by Landis and Koch.¹² Values less than 0.00 were considered as poor reliability/agreement, 0.01 to 0.20 as slight reliability/agreement, 0.21 to 0.40 as fair reliability/agreement, 0.41 to 0.60 as moderate reliability/agreement, 0.61 to 0.80 as substantial reliability/agreement, 0.81 to 0.99 as almost perfect reliability/agreement, and 1.00 as perfect reliability/agreement. The overall concordance rate for each parameter across all cases was calculated and expressed as a percentage. A *P*-value < 0.05 was considered statistically significant.

Ethics approval

All biopsies were obtained from patients enrolled in the SPARC-IBD multicenter cohort, a component of the IBD Plexus of the Crohn's and Colitis Foundation. SPARC-IBD data are available upon approved application to the Crohn's and Colitis Foundation IBD Plexus (<https://www.crohnscolitisfoundation.org/ibd-plexus>).

Results

Site of biopsies and endoscopic scores

Of the 37 biopsies, 11 (30%) were from the rectum, 15 (40%) from the sigmoid colon, four (11%) from the descending colon, three (8%) from the ascending colon, and four (11%) from the cecum. Nine biopsies (24%) corresponded to an MES of 0, 19 (51%) to an MES of 1, six (16%) to an MES of

Table 1. Interrater reliability for parameters assessed in grading previously treated ulcerative colitis activity in the cohort of 37 biopsies using the Nancy histologic index

Item	Interrater reliability, ICC	
	Review 1	Review 2
	ICC (95% CI)	ICC (95% CI)
Overall NHI Grade	0.79 (0.70–0.87)	0.78 (0.69–0.86)
Grade 0	0.74 (0.64–0.83)	0.75 (0.65–0.84)
Grade 1	0.75 (0.66–0.84)	0.79 (0.71–0.87)
Grade 2	0.24 (0.15–0.37)	0.23 (0.14–0.36)
Grade 3	0.42 (0.30–0.56)	0.47 (0.35–0.61)
Grade 4	0.41 (0.29–0.55)	0.47 (0.35–0.61)
Grades 2, 3 and 4 combined (active disease)	0.76 (0.66–0.85)	0.76 (0.66–0.85)

CI, confidence interval; ICC, intraclass correlation coefficient; NHI, Nancy histologic index.

2, and the remaining three (8%) to an MES of 3. Based on majority grading, the distribution of cases at the first reading was as follows: 21 out of 37 cases (56.7%) were classified as Grade 0, one case (3.7%) as Grade 1, five cases (13.5%) as Grade 2, seven cases (18.9%) as Grade 3, and three cases (8.1%) as Grade 4. At the second reading, the distribution was 21 cases (56.7%) as Grade 0, one case (3.7%) as Grade 1, six cases (16.0%) as Grade 2, five cases (13.5%) as Grade 3, and four cases (10.8%) as Grade 4.

IRR of NHI

The IRR for the overall NHI among the 12 pathologists was substantial at both reviews [Review 1: ICC = 0.79, 95% confidence interval (CI): 0.70–0.87; Review 2: ICC = 0.78, 95% CI: 0.69–0.86]. However, there was considerable variability in IRR among the individual NHI grades. Grades 0 and 1 showed substantial IRR at both reviews ($P < 0.001$). Grades 3 and 4 demonstrated moderate IRR at both reviews ($P < 0.001$). Grade 2 had only fair IRR at both reviews ($P < 0.001$) (Table 1). When Grades 2, 3, and 4 were combined, the IRR remained substantial at both reviews (ICC = 0.76, 95% CI: 0.66–0.85; $P < 0.001$).

Intra-rater agreements for NHI and its components

The mean intra-rater agreement for NHI and its components, assessed using unweighted Cohen's kappa, is summarized in Table 2. Grade 2 had the lowest intra-rater agreement (fair). Grade 4 showed moderate intra-rater agreement among the participating pathologists. Figure 1 illustrates the concordant

and discordant cases for each NHI component across both reviews, highlighting inconsistency, particularly in assigning Grades 2 and 4. H&E-stained colorectal biopsies from representative cases with the least concordance in Grades 2 and 4 are shown in Figure 2a–d.

IRR and intra-rater agreements for crypt distortion and Paneth cell metaplasia

The IRR for crypt distortion was substantial at both reviews (Review 1: ICC = 0.64 (95% CI: 0.52–0.75, $P < 0.001$) and Review 2: ICC = 0.68 (95% CI: 0.57–0.79, $P < 0.001$)). The IRR for Paneth cell metaplasia was moderate (Review 1: ICC = 0.60 (95% CI: 0.49–0.73, $P < 0.001$) and Review 2: ICC = 0.51 (95% CI: 0.39–0.65, $P < 0.001$)). The mean intra-rater agreement for crypt distortion and Paneth cell metaplasia was substantial, with a Cohen's kappa = 0.70 (range: 0.45–0.95) and 0.62 (range: 0.04–1.0), respectively.

Discussion

We observed substantial IRR for the NHI among 12 practicing pathologists with subspecialty training in gastrointestinal pathology, both before (ICC = 0.79, 95% CI: 0.70–0.87, $P < 0.001$) and after (ICC = 0.78, 95% CI: 0.69–0.86, $P < 0.001$) the implementation of a brief online tutorial on the NHI. When analyzing individual NHI components, we found that Grade 1 exhibited the highest IRR at both assessments, while Grade 2 showed the lowest IRR, with minimal improvement between reviews. Grades 3 and 4 had intermediate IRR

Table 2. Mean intra-rater agreement for parameters assessed in grading previously treated ulcerative colitis activity in the cohort of 37 biopsies using the Nancy index

Feature/item	Intra-rater agreement, Cohen's kappa (range)	Interpretation
Overall NHI Grade	0.57 (0.24–0.80)	Moderate agreement
Grade 0	0.81 (0.52–1)	Substantial agreement
Grade 1	0.74 (0.52–1.0).	Substantial agreement
Grade 2	0.31 (–0.08–0.77)	Fair agreement
Grade 3	0.51 (0.08–0.91)	Moderate agreement
Grade 4	0.59 (0.2–1.0)	Moderate agreement
Combined Grades 2, 3 and 4 (active disease)	0.75 (0.37–1.0)	Substantial agreement

NHI, Nancy histologic index.

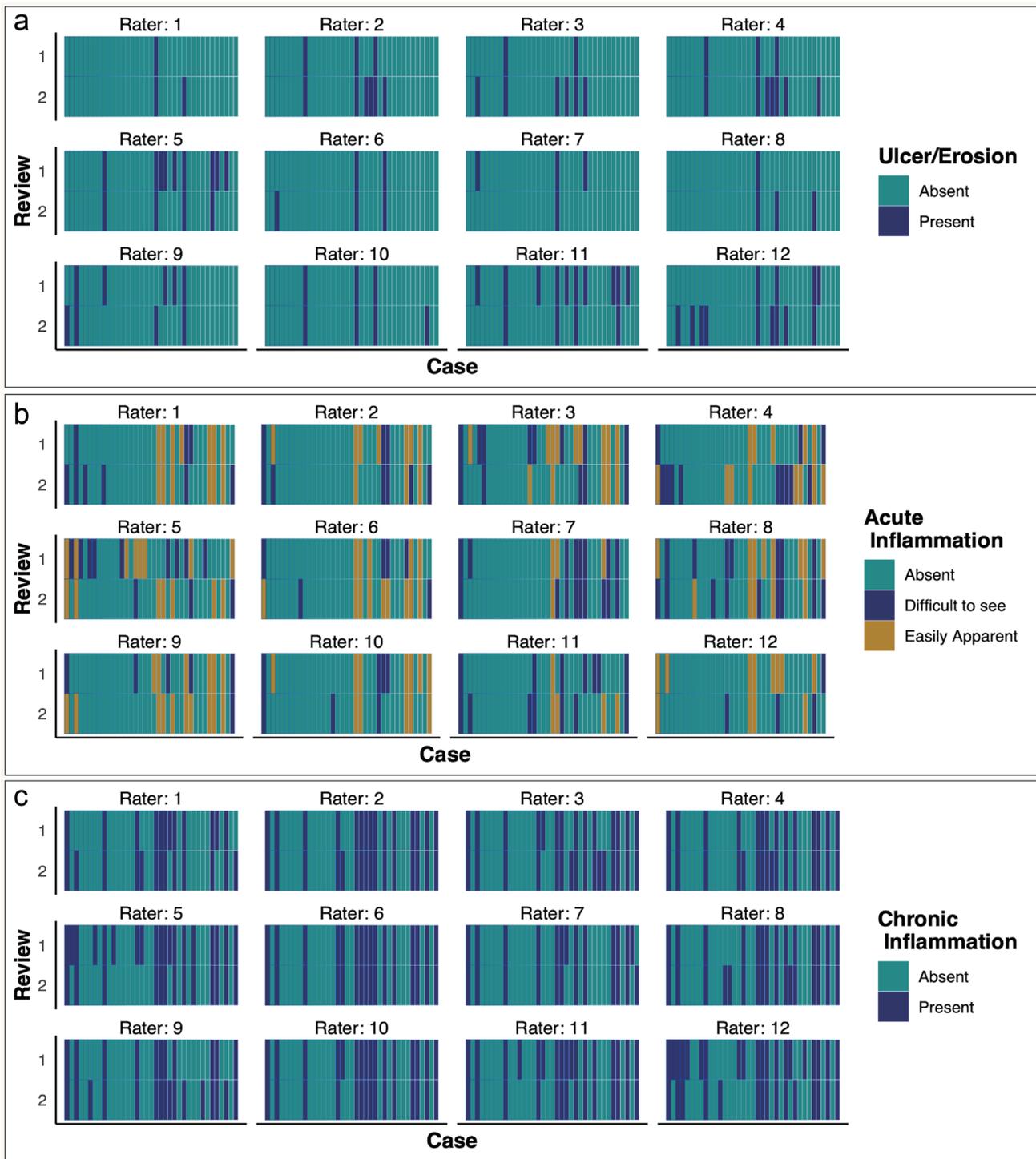


Fig. 1. Concordance by case and review for the Nancy histologic index components. Concordance across 37 cases is shown for each pathologist at both reviews for the three components of the Nancy histologic index: ulceration (a), acute inflammation (b), and chronic inflammation (c). Each colored bar represents an individual case. Considerable discordance was observed for cases classified as Grade 2 (blue bars in b) and Grade 4 (blue bars in a) at both reviews.

values. Notably, combining Grades 2, 3, and 4 (i.e., active disease) yielded substantial IRR.

While Marchal-Bressenot *et al.*⁸ demonstrated near-perfect IRR for the NHI (ICC = 0.88, 95% CI: 0.82–0.92) and

its component items—except for chronic inflammation (ICC = 0.63, 95% CI: 0.33–0.70)—discrepancies in other studies, including ours, highlight some challenges of using this index. Similar to our findings, Jairath *et al.*¹³ reported substantial

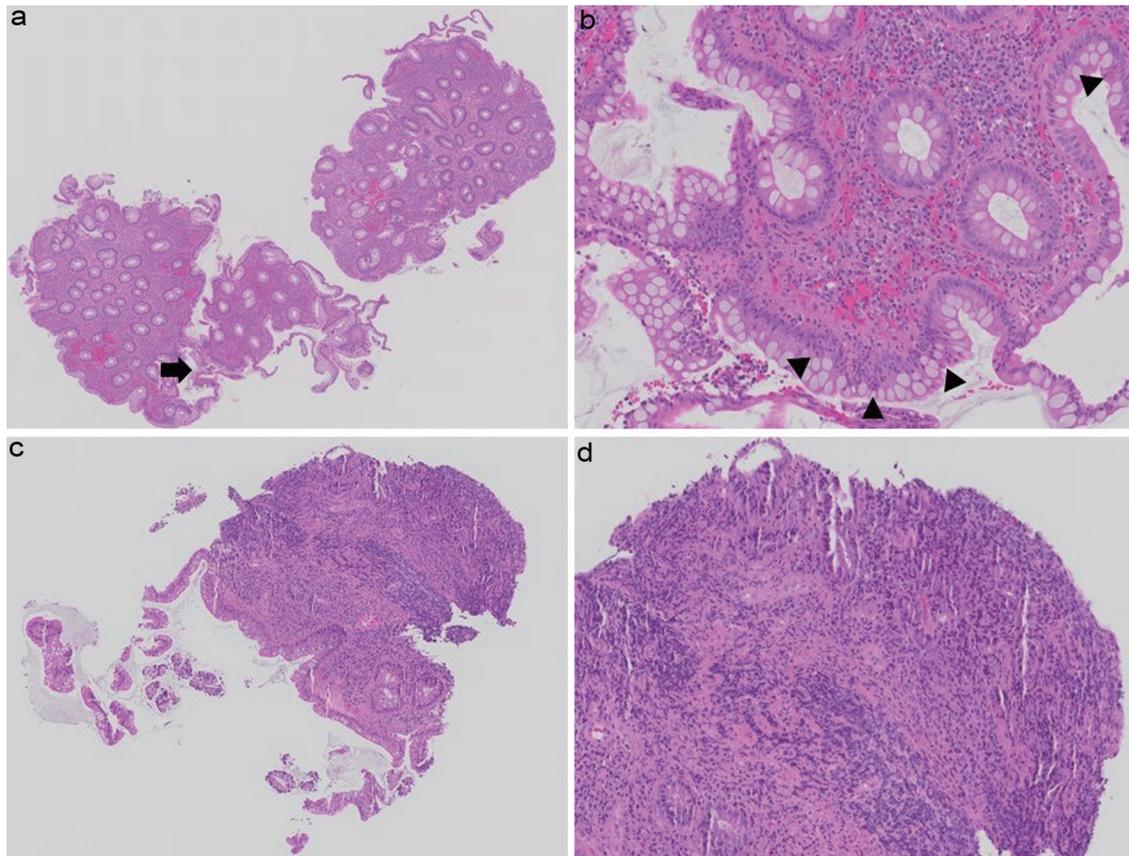


Fig. 2. Discrepancy in Nancy index grading of hematoxylin and eosin-stained biopsy images by pathologists. Sigmoid colon biopsy from a patient with treated ulcerative colitis showing chronic inflammation in the lamina propria and a focal area (black arrow in a) with intraepithelial and lamina propria neutrophils (arrowheads in b). At the first review, 8/12 pathologists assigned Grade 2, 3/12 assigned Grade 3, and 1/12 assigned Grade 1. At the second review, 7/12 pathologists assigned Grade 2, and 5/12 assigned Grade 3. Of note, 11/12 pathologists at review 1 and 12/12 at review 2 indicated that this biopsy had active disease. Rectal biopsy from a patient with treated ulcerative colitis showing chronic inflammatory infiltrate in the lamina propria (c) and a possible ulceration with granulation tissue-like appearance (d). At the first review, 7/12 pathologists assigned Grade 4, 4/12 assigned Grade 1, and 1/12 assigned Grade 2. At the second review, 7/12 pathologists assigned Grade 4, 2/12 assigned Grade 2, and Grades 0, 1, and 3 were each assigned by 1/12 pathologists. Notably, 9/12 pathologists at review 1 and 10/12 at review 2 indicated that this biopsy had active disease. (Hematoxylin-eosin stain; original magnifications: 3.2x in a, 20x in b, 4x in c, and 20x in d).

IRR for final NHI grades (ICC = 0.80, 95% CI: 0.73–0.85) among four pathologists with expertise in inflammatory bowel disease. Le *et al.*¹⁴ also reported substantial IRR for final NHI grades (ICC = 0.70, 95% CI: 0.50–0.82) between two pathologists, including a pathologist-in-training, but noted higher discordance for Grades 1 and 4. Arkteg *et al.*¹⁵ reported substantial IRR for the presence of acute inflammation (ICC = 0.79, 95% CI: 0.64–0.88). However, their study did not provide IRR values for Grades 2 and 3 individually. Like our findings, they reported lower IRR for Grade 4 (ICC = –0.04, 95% CI: –0.74 to 0.41), although it remains unclear whether biopsies with erosions were included in this group. Notably, they also observed low IRR for chronic inflammation (ICC = 0.42, 95% CI: 0.02–0.67). Discrepancies across studies may be attributed to factors such as case selection, use of glass slides versus digital images, practice settings, and the level of experience among participating pathologists. These variations underscore the challenges of assessing certain NHI components.

Subjectivity in distinguishing Grades 2 and 3—specifically, the need to identify a few or rare neutrophils (often difficult to visualize) versus multiple, easily visible clusters—may have contributed to the lower IRR. Neutrophils are not normally present in the intestinal mucosa, and the

threshold of neutrophilic inflammation that increases the risk of adverse outcomes has yet to be clearly defined. The Geboes Score considers both lamina propria and epithelial neutrophils, providing a quantitative evaluation of the latter. Similarly, the Robarts Histologic Index, which is largely derived from the Geboes Score, assesses neutrophils in both compartments. These indices, unlike the NHI, also distinguish between erosions and ulcers. In this study, erosions were categorized as NHI Grade 4. Encouragingly, the substantial IRR for active disease (Grades 2, 3, and 4) in our study underscores the NHI’s clinical utility. However, refining the criteria for these grades will be essential for reducing inter-observer variability and enabling more accurate monitoring of treatment endpoints. This may include developing more precise definitions for the amount of acute inflammation that qualifies as Grade 2 or 3, and clarifying the classification of erosions—an issue currently unaddressed by the NHI. It may also be important to specify whether neutrophils exclusively located in the lamina propria should be considered Grade 2. Notably, the two additional features evaluated in our study—crypt architectural distortion and Paneth cell metaplasia—had moderate IRR in both reviews. Although these parameters are not part of the NHI, they are routinely used in clinical practice as markers of chronic

mucosal injury and are included in other scoring systems, such as the Geboes Score.

More recently, artificial intelligence (AI)-powered algorithms have been applied to UC datasets to assist in the histologic grading of biopsies.^{16–19} Najdawi *et al.*¹⁶ used convolutional neural networks to segment tissue and classify cells on whole-slide H&E-stained biopsies to generate NHI predictions. Their AI model showed strong correlation with increasing NHI scores ($\rho = 0.90$, $P < 0.001$) and reliably distinguished between different grades based on the proportion of epithelium with neutrophilic inflammation, the count and density of neutrophils in the epithelium, and the presence of ulcers or combinations thereof ($\rho = 0.83–0.90$, all $P < 0.001$). Peyrin-Biroulet *et al.*¹⁷ employed four artificial neural networks to recognize cell types and assign NHI grades. They found that the AI-based grading was reproducible and comparable in performance (ICC = 87.2%) to that of four expert histopathologists (ICC = 89.3%). The PICaSSO Histologic Remission Index, a recently introduced simplified scoring system, focuses on the presence or absence of neutrophils in the epithelium (surface and crypt) and lamina propria. This index has shown stronger correlation with endoscopic activity compared to other histologic indices, including the NHI, and exhibits minimal inter-rater variability.¹⁸ It has also been validated using an AI model, which accurately and reliably predicted PICaSSO Histologic Remission Index.

While our study provides valuable insight into the reproducibility of histologic assessments of colorectal biopsies from treated UC patients using the NHI, several limitations should be acknowledged. The small sample size and uneven distribution of biopsy sites may have led to an underestimation of the ICC.¹⁹ Additionally, this study focused on reproducibility among academic gastrointestinal pathologists, so results may not be fully generalizable to real-world practice settings where levels of expertise may vary. Notably, only two of the reviewing pathologists had prior experience with a modified version of the NHI. Additionally, variations in staining quality and image artifacts may have contributed to interpretation differences.

Conclusions

Our study revealed substantial IRR for active disease (Grades 2, 3, and 4) among 12 pathologists, which underscores the clinical utility of the NHI in the assessment of colorectal biopsies from treated UC patients. However, refinement of the criteria for Grades 2, 3, and 4 may be required to improve reproducibility and enable more accurate monitoring of treatment outcomes in UC, especially as histologic remission is an evolving therapeutic endpoint.

Acknowledgments

The results published here are, in whole or in part, based on data from the inflammatory bowel disease (IBD) Plexus program of the Crohn's and Colitis Foundation. The Study of a Prospective Adult Research Cohort with Inflammatory Bowel Disease (SPARC IBD) is a component of the Crohn's & Colitis Foundation's IBD Plexus data exchange platform. SPARC IBD enrolls patients with a new or established diagnosis of IBD from sites across the United States and links data collected from electronic health records and study-specific case report forms. Patients also provide blood, stool, and biopsy samples at designated time points during follow-up. The design and implementation of the SPARC IBD cohort have been previously described.

Funding

This study was completed without financial support.

Conflict of interest

Dr. Parakkal Deepak is supported by a Junior Faculty Development Award from the American College of Gastroenterology and the inflammatory bowel disease (IBD) Plexus program of the Crohn's & Colitis Foundation. Three of the authors—Dr. Hanlin L. Wang, Dr. Zhaohai Yang, and Dr. Xiuli Liu—are Editorial Board Members, and one author, Dr. Xuchen Zhang, is the Associate Editor of the *Journal of Clinical and Translational Pathology* since May 2021. The authors declare no other conflicts of interest.

Author contributions

Study conception, design, statistical data interpretation (XL), data curation (SM, PD, KDS, JNL), statistical analysis and original draft preparation (KDS, JNL), and histopathologic review (DA, SJB, KB, AGD, AKE, RSG, XG, HL, JML, NS, HLW, ZY, XZ). All authors have reviewed and approved the final version of the manuscript.

SPARC-IBD investigators

Richa Shukla: Baylor College of Medicine; Themistocles Dasopoulos: Baylor University Medical Center; Scott B. Snapper, Joshua R. Korzenik: Brigham & Women's Hospital; Matthew Bohm: Indiana University; Laura Raffals: Mayo Clinic; Poonam Beniwal-Patel: Medical College of Wisconsin; David Hudesman: NYU Langone Medical Center; Mazer Ally, Gauree Konijeti, Rebecca Matro: Scripps Healthcare; Sheldon Lidofsky: Brown University; Kirk Russ: University of Alabama; Loren Brook: University of Cincinnati Medical Center; Joel Pekow: University of Chicago; Raymond Cross: University of Maryland; Shrinivas Bishu: University of Michigan; Meenakshi Bewtra, James D Lewis: University of Pennsylvania; Richard Duerr: University of Pittsburgh; Sumona Saha, Freddy Caldera: University of Wisconsin; Elizabeth Scoville: Vanderbilt University Medical Center; Parakkal Deepak: Washington University School of Medicine.

Ethical statement

This study was approved by the Ethics Committee of Washington University (Approval No. 202206060) and was conducted in accordance with the Declaration of Helsinki (as revised in 2024). As the data were obtained from an existing database, the requirement for informed consent was waived.

Data sharing statement

The dataset used in support of the findings of this study are included within the article.

References

- [1] Gupta RB, Harpaz N, Itzkowitz S, Hossain S, Matula S, Kornbluth A, *et al.* Histologic inflammation is a risk factor for progression to colorectal neoplasia in ulcerative colitis: a cohort study. *Gastroenterology* 2007;133(4):1099–1105. doi:10.1053/j.gastro.2007.08.001, PMID:17919486.
- [2] Shehab M, Al Akram S, Hassan A, Alrashed F, Jairath V, Bessissow T. Histological Disease Activity as Predictor of Clinical Relapse, Hospitalization, and Surgery in Inflammatory Bowel Disease: Systematic Review and Meta-Analysis. *Inflamm Bowel Dis* 2024;30(4):563–572. doi:10.1093/ibd/izad119, PMID:37541185.
- [3] Yoon H, Jangi S, Dulai PS, Boland BS, Prokop LJ, Jairath V, *et al.* Incremental Benefit of Achieving Endoscopic and Histologic Remission in Patients

- With Ulcerative Colitis: A Systematic Review and Meta-Analysis. *Gastroenterology* 2020;159(4):1262–1275.e7. doi:10.1053/j.gastro.2020.06.043, PMID:32585306.
- [4] Turner D, Ricciuto A, Lewis A, D’Amico F, Dhaliwal J, Griffiths AM, *et al*. STRIDE-II: An Update on the Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE) Initiative of the International Organization for the Study of IBD (IOIBD): Determining Therapeutic Goals for Treat-to-Target strategies in IBD. *Gastroenterology* 2021;160(5):1570–1583. doi:10.1053/j.gastro.2020.12.031, PMID:33359090.
- [5] Pai RK, D’Haens G, Kobayashi T, Sands BE, Travis S, Jairath V, *et al*. Histologic assessments in ulcerative colitis: the evidence behind a new endpoint in clinical trials. *Expert Rev Gastroenterol Hepatol* 2024;18(1-3):73–87. doi:10.1080/17474124.2024.2326838, PMID:38509826.
- [6] Ma C, Sedano R, Almradi A, Vande Casteele N, Parker CE, Guizzetti L, *et al*. An International Consensus to Standardize Integration of Histopathology in Ulcerative Colitis Clinical Trials. *Gastroenterology* 2021;160(7):2291–2302. doi:10.1053/j.gastro.2021.02.035, PMID:33610533.
- [7] Nardone OM, Iacucci M, Villanacci V, Peyrin-Biroulet L, Ghosh S, Danese S, *et al*. Real-world use of endoscopic and histological indices in ulcerative colitis: Results of a global survey. *United European Gastroenterol J* 2023;11(6):514–519. doi:10.1002/ueg2.12423, PMID:37323091.
- [8] Marchal-Bressenot A, Salleron J, Boulagnon-Rombi C, Bastien C, Cahn V, Cadiot G, *et al*. Development and validation of the Nancy histological index for UC. *Gut* 2017;66(1):43–49. doi:10.1136/gutjnl-2015-310187, PMID:26464414.
- [9] Marchal-Bressenot A, Scherl A, Salleron J, Peyrin-Biroulet L. A practical guide to assess the Nancy histological index for UC. *Gut* 2016;65(11):1919–1920. doi:10.1136/gutjnl-2016-312722, PMID:27566129.
- [10] Magro F, Doherty G, Peyrin-Biroulet L, Svrcek M, Borralho P, Walsh A, *et al*. ECCO Position Paper: Harmonization of the Approach to Ulcerative Colitis Histopathology. *J Crohns Colitis* 2020;14(11):1503–1511. doi:10.1093/ecco-jcc/jjaa110, PMID:32504534.
- [11] 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):1625–1629. doi:10.1056/NEJM198712243172603, PMID:3317057.
- [12] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159–174. PMID:843571.
- [13] Jairath V, Peyrin-Biroulet L, Zou G, Mosli M, Vande Casteele N, Pai RK, *et al*. Responsiveness of histological disease activity indices in ulcerative colitis: a post hoc analysis using data from the TOUCHSTONE randomised controlled trial. *Gut* 2019;68(7):1162–1168. doi:10.1136/gutjnl-2018-316702, PMID:30076171.
- [14] Le HD, Pflaum T, Labrenz J, Sari S, Bretschneider F, Tran F, *et al*. Interobserver Reliability of the Nancy Index for Ulcerative Colitis: An Assessment of the Practicability and Ease of Use in a Single-Centre Real-World Setting. *J Crohns Colitis* 2023;17(3):389–395. doi:10.1093/ecco-jcc/jjac146, PMID:36282973.
- [15] Arkteg CB, Wergeland Sørbye S, Buhl Riis L, Dalen SM, Florholmen J, Goll R. Real-life evaluation of histologic scores for Ulcerative Colitis in remission. *PLoS One* 2021;16(3):e0248224. doi:10.1371/journal.pone.0248224, PMID:33684168.
- [16] Najdawi F, Sucipto K, Mistry P, Hennek S, Jayson CKB, Lin M, *et al*. Artificial Intelligence Enables Quantitative Assessment of Ulcerative Colitis Histology. *Mod Pathol* 2023;36(6):100124. doi:10.1016/j.modpat.2023.100124, PMID:36841434.
- [17] Peyrin-Biroulet L, Adsul S, Stancati A, Dehmeshki J, Kubassova O. An artificial intelligence-driven scoring system to measure histological disease activity in ulcerative colitis. *United European Gastroenterol J* 2024;12(8):1028–1033. doi:10.1002/ueg2.12562, PMID:38590110.
- [18] Gui X, Bazarova A, Del Amor R, Vieth M, de Hertogh G, Villanacci V, *et al*. PICaSSO Histologic Remission Index (PHRI) in ulcerative colitis: development of a novel simplified histological score for monitoring mucosal healing and predicting clinical outcomes and its applicability in an artificial intelligence system. *Gut* 2022;71(5):889–898. doi:10.1136/gutjnl-2021-326376, PMID:35173041.
- [19] Mehta S, Bastero-Caballero RF, Sun Y, Zhu R, Murphy DK, Hardas B, *et al*. Performance of intraclass correlation coefficient (ICC) as a reliability index under various distributions in scale reliability studies. *Stat Med* 2018;37(18):2734–2752. doi:10.1002/sim.7679, PMID:29707825.